Europäisches
Patentamt

European Patent
Office

Office européen
des brevets

#2½
D.g.J.
10-9-91

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen
mit der ursprünglich eingereichten Fas-
sung der auf dem nächsten Blatt be-
zeichneten europäischen Patentanmel-
dung überein.

The attached documents are exact
copies of the European patent applica-
tion described on the following page, as
originally filed.

Les documents fixés à cette attestation
sont conformes à la version initialement
déposée de la demande de brevet euro-
péen spécifiée à la page suivante.

Den Haag, den
The Hague,
La Haye, le

2 0. 06. 90

Der Präsident des Europäischen Patentamts
Im Auftrag
For the President of the European Patent Office
Le Président de l'Office européen des brevets
p. o.

L.J.M. van der Heijdt

Patentanmeldung Nr.
Patent application no.   89202117.1
Demande de brevet n°

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: *89202117.1*
Demande n°:

Anmeldetag:
Date of filing: *11.08.1989*
Date de dépôt:

Anmelder:
Applicant(s): *GIST-BROCADES*
Demandeur(s):

Delft - The Netherlands

Bezeichnung der Erfindung:
Title of the invention: *Efficient production of mutant proteases.*
Titre de l'invention:

In Anspruch genommene Priorität(en)
Priority(ies) claimed
Priorité(s) revendiquée(s)

Staat:	Tag:	Aktenzeichen:
State: - - -	Date: - - -	File no. - - -
Pays:	Date:	Numéro de dépôt:

Internationale Patentklassifikation:
International Patent Classification: *C 12 N 15/57, C 12 N 9/54,C 12 N 15/75,C 12 N 1/2i//,*
Classification internationale des brevets:
(C 12 N 1/21 , C 12 R 1: 07)

Am Anmeldetag benannte Vertragsstaaten:
Contracting states designated at the date of filing: AT-BE-CH-LI-DE-ES-FR-GB-GR-IT-LU-NL-SE
Etats contractants désignés lors du dépôt:

Bemerkungen:
Remarks:
Remarques:

EFFICIENT PRODUCTION OF MUTANT PROTEASES

INTRODUCTION

5   ## Technical Field

The present invention relates to a method for the production of mutant proteins using recombinant DNA techniques. More specifically, the invention relates to the
10 efficient production of protein engineered Bacillus proteases using a homologous Bacillus host strain, from which the capacity to produce the corresponding unmodified protease has been eliminated.

15   ## Background of the Invention

Bacilli are widely used for the production of industrially important enzymes such as α-amylases, neutral proteases and alkaline (or serine) proteases, see, for
20 example, Debabov: "The industrial use of Bacilli" in: The Molecular Biology of Bacilli, Acad. Press, New York, 1982.
The main advantages of using Bacilli are the large quantities of protein secreted, their lack of production of hazardous substances and their long history of safe
25 industrial use. Several species of Bacilli have therefor been approved for the production of proteins destined for food applications.
The use of Bacillus subtilis as an expression host was demonstrated by Palva et al., Gene 19 (1982) 81-87,
30 describing the expression of the α-amylase gene from B. amyloliquefaciens in B. subtilis. Other references describing heterologous gene expression in B. subtilis are, e.g., Fahnestock and Fisher, J. Bacteriol. 165 (1986) 796-804, who showed the expression of the Staphylococcus aureus
35 Protein A gene; Schein et al., Biotechnology 4 (1986) 719-725, disclosing the expression of human interferon-α2; and Wang et al., Gene 69 (1988) 39-47, showing the expression

and secretion of human atrial natriuretic α-factor. To ensure the proper secretion of these heterologous proteins the former two references used the α-amylase signal sequence and the latter reference the signal sequence of the B. subtilis subtilisin gene (aprE) fused to the gene to be expressed.

A major problem of Bacillus strains as production hosts for recombinant proteins is that they produce and secrete a variety of proteases which tend to degrade the heterologous proteins produced. See, e.g., Old and Primrose, "Principles of Gene Manipulation", 3rd ed. (1985) 289, Blackwell, Oxford.

In International Patent Application (PCT) WO 89/04866 the importance of extracellular proteases is pointed out; about 90% of the proteolytic activity is attributed to the neutral metalloprotease (npr) and the serine protease (apr).

Various solutions have been proposed to overcome this degradation problem.

Kawamura and Doi, J. Bacteriol. 160 (1984) 442-444, disclose a B. subtilis double mutant deficient in both extracellular neutral and alkaline protease. First B. subtilis DB100 was made neutral protease negative by transferring the nprR2 and the nprE18 mutations from B. subtilis NT18. A 178 bp fragment containing a part of the apr gene was then deleted by gene conversion.

Fahnestock and Fisher, Appl. Environ. Microbiol. 53 (1987) 379-384, describe the construction of an apr B. subtilis strain starting from a npr strain. They used a plasmid vector containing the apr gene and introduced the cat gene both as an inactivator and a selectable marker.

Sloma et al., J. Bacteriol. 170 (1988) 5557-5563, detected and sequenced a third extracellular protease gene (epr) from B. subtilis. Partial deletion of this gene showed that it contributed only marginally to the extracellular protease activity.

However, none of the referenced B. subtilis strains was completely devoid of extracellular proteolytic activity.

In several publications the production of mutant
5 proteases is described by Bacillus subtilis host strains which are incapable of expressing the wild-type protease gene, thus avoiding mixtures of mutant and wild-type proteases.

In EP-A-0246678 a B. subtilis strain is disclosed
10 which is incapable of secreting enzymatically active subtilisin or neutral protease, but which expresses heterologous protease genes isolated or derived from B. amyloliquefaciens. Site-specific saturation mutagenesis was performed on the B. amyloliquefaciens gene and the mutants
15 were expressed. The negative background of the B. subtilis host strains was obtained by partial deletion of either or both of the original protease genes, or by N-methyl-N'-nitro-N-nitrosoguanidine (NTG) mutagenesis. It was further stressed that the B. subtilis host described is normally
20 sporulating and that asporogenous mutants are unsatisfactory for the production of (recombinant) proteins.

In WO 86/01825 Bacillus strains are described, in particular B. subtilis, with reduced extracellular protease levels. The alkaline protease gene is inactivated by in
25 vitro insertion of a gene coding for a protein which confers a phenotypic trait on the microorganism. The plasmid vector containing the inactivated protease gene is transformed to B. subtilis and the inactivated gene replaces the native chromosomal gene by homologous recombination. No expression
30 and secretion of heterologous proteins is suggested.

WO 88/8033 discloses subtilisin analogs from subtilisin Carlsberg, subtilisin DY, subtilisin BPN', an apr subtilisin from B. subtilis and a subtilisin from B. mesentericus. These mutants are obtained by replacing one or
35 more amino acids in or near the calcium binding site with negatively charged amino acids (e.g. Asp or Glu). The mutants are expressed in B. subtilis.

WO 89/04866 discloses a B. subtilis strain with low secreted protease activity, which was made by a spoOH mutation in an apr⁻npr⁻ double mutant. The obtained triple mutant apr⁻npr⁻spoOH⁻ showed very low secreted protease

5 activity.

The references discussed above describe the expression of homologous and heterologous genes in B. subtilis and the construction of protease-negative B. subtilis strains. It has appeared that protease-negative strains

10 made by partial deletion of the protease gene or by insertion of another gene in the protease gene show several drawbacks, e.g.:

- the original gene may be reactivated. If a gene which is homologous to the inactivated chromosomal gene is

15 introduced into the cell (on a plasmid), homologous recombination can lead to reactivation of the gene.

- the original gene, although inactivated, may lead to the production of partial expression products, when the promoter region is still active. This is disadvantageous in

20 terms of metabolic efficiency.

- the introduction of unwanted phenotypical traits, e.g. antibiotic resistance, depending on the inserted gene.

- plasmid instability, due to homology between a genomic sequence and a part of the inserted plasmid.

25 It is therefore desirable to delete all sequences with possible homology between the chromosome and the plasmid.

There are also other drawbacks for the production of alkaline proteases or other enzymes in B. subtilis. Although

30 Wells et al., Nucleic Acids Res. 11, (1983) 7911-7925 shows that expression of B. amyloliquefaciens subtilisin BPN' in B. subtilis is possible with its own transcription signals, Jacobs et al., Nucleic. Acids Res. 13 (1985) 8913-8926 shows that such a strategy is not generally applicable. It

35 was demonstrated that the 5' region of subtilisin Carlsberg derived from B. licheniformis did not contain functional signals for transcription in B. subtilis. To achieve

considerable production it was needed to insert a B. subtilis promoter in front of the gene in question. It is also suggested that not only transcriptional but also translational, secretion or maturation processes may occur
5 less efficient in heterologous hosts, due to a not perfect Shine Dalgarno sequence or even incompatibility of the gene product with the heterologous host.

The problems mentioned above can be avoided when the product is synthesized in a homologous Bacillus strain, of
10 which high production capabilities are proven, with high efficiency resulting in high production levels.

SUMMARY OF THE INVENTION

15 Transformed alkalophilic Bacillus strains are provided for the efficient production strains of mutant Bacillus proteases. The new strains have lost their ability to express their protease encoding genes originally present.

In a preferred embodiment mutant proteases are
20 produced which show close homology with the proteases originally produced by the corresponding strain.

Preferred strains are Bacillus novo species PB92, derivatives thereof and closely related strains, transformed by a mutated protease encoding gene. Said gene shows at
25 least 30%, preferably at least 50% and more preferably at least 80% homology with the wild-type gene of Bacillus PB92 encoding serine protease. The gene is preferably derived from a wild-type gene of an alkalophilic Bacillus strain.

Also provided are methods of constructing the new
30 transformed strains, vectors used thereby and a method of producing mutant proteases.

BRIEF DESCRIPTION OF THE DRAWINGS

35 Figure 1 shows the construction of plasmid pM58 Δ containing parts of the 5' and 3' flanking regions of the alkaline protease gene from Bacillus novo sp. PB92.

Figure 2 schematically shows the introduction of the temperature-sensitive origin of replication from plasmid pE194neo-1 into plasmid pM58Δ, giving the plasmid pEM58Δ.

Figure 3 shows the introduction by homologous recombination, of the plasmid pEM58Δ into the 5' flanking region of the protease gene of Bacillus strain PBT110.

Figure 4 shows a simplified restriction map of the regions surrounding the deleted protease gene after illegitimate recombination. Two strains were obtained, PBT125 and PBT126.

Figure 5 nucleotide sequence of the HindIII fragments of PBT125 (A) and PBT126 (B).

Figure 6 schematic representation of the introduction of the protease gene into M13 mp18.

Figure 7 shows the nucleotide sequence of plasmid pBHA-1.

Figure 8 shows the introduction of the protease gene of Bacillus strain PBT110 in plasmid pBHA1.

Figure 9 shows the reorientation of the F1 ori containing BamHI fragment of pBHA1-MXL giving plasmid pBHAR-MXL.

Figure 10 shows the subcloning of the 3' sequence of the high alkaline protease into plasmid pBHAR-MXL, giving plasmid pBHARB-MXL.

Figure 11 shows the deletion of the E. coli sequence from plasmid pBHARB-MXL M216Q giving plasmid pBHB-MXL M216Q.

Figure 12 shows the reorientation of the neomycin resistance gene in plasmid pE194neo-1 giving plasmid pE194neo-3.

Figure 13 schematically shows the introduction of the temperature-sensitive origin of replication from plasmid pE194neo-3 into plasmid pBHB-MXL M216Q giving plasmid pEN$_3$Q.

Figure 14 schematically shows the construction of strain Bacillus strain PEP111.

Figure 15 schematically shows the construction of strain Bacillus strain PEP211.

DETAILED DESCRIPTION OF THE INVENTION

For the production of alkaline proteases preferably alkalophilic Bacillus strains are used as host cells. For
5 large scale production industrial strains are a need. They originate from organisms which may be isolated from the soil or are available from depositories or other sources and are obtained by genetic modification of such Bacillus strains. Industrial Bacillus strains are defined in EP-A-
10 134048. They are characterized by being resistant to genetic exchange, such as phage infection or transformation. The strains are stable and may or may not be capable of spore formation. They are usually prototrophic and modified to provide for high yields of endogenous protein products, such
15 as the enzymes alpha-amylase and various proteases. The yield of an endogenous protein product obtained in an industrial production process can amount to at least 5 g/l (0.5% w/v). Industrial strains also secrete DNases, which result in degradation of DNA in the medium, providing for
20 protection against genetic exchange.

An additional advantage of using alkalophilic Bacilli as production strains is the minimal risk of contamination with other microorganisms, because of the alkaline pH during fermentation.
25 According to the present invention a method is disclosed for the deletion of a Bacillus protease encoding gene from the genome in such a way that reversion becomes impossible. Said deletion may be a partial deletion but more preferably the gene is completely deleted. The protease
30 negative strain obtained by this method is used for the production of homologous mutant proteases after introduction of an expression vector carrying the gene encoding this mutant protease. It is, however, understood that such a protease negative strain will also be advantageous for the
35 expression of other proteins.

It has been found that by using an industrial protease producing Bacillus strain as a host for the

production of the mutants of this protease, the high efficiency of the production of the original protease is transferred to the mutant protease. This results in a surprisingly superior production efficiency compared to laboratory strains. One chromosomal copy of the mutated gene gives upon expression and secretion in an industrial strain a yield even higher than 50 plasmid copies containing the mutated gene, in a laboratory strain. Moreover, as for the industrial strain a production process including fermentation, recovery and formulation protocols is available the invention that industrial strains are particularly useful for mutant protease production has the advantage of not having to develop a production process as would be the case if a new production host strain was chosen.

It is an aspect of this invention that mutant proteases can be produced efficiently by a homologous _Bacillus_ strain by way of exchanging the chromosomal gene or a part thereof with the corresponding mutated gene. In this way the production capacity for the wild-type protease is eliminated while at the same time introducing production capacity for a mutant protease.

As a host cell for the expression of mutated genes encoding modified or so-called protein engineered proteins it is preferable to use cells in which the genes are structurally expressed at high level.

The present invention also shows (Table 1) that to get a sufficient production of the high alkaline protease derived from _Bacillus_ PB92 in _B_. _subtilis_, it was needed to insert a promoter in front of the gene which is active in _B_. _subtilis_. If enzymes derived from other _Bacilli_ are produced in _B_. _subtilis_ insertion of a promoter active in _B_. _subtilis_ or other expression signals may be necessary, which requires an extra step.

High alkaline protease producing bacilli are taxonomically not well classified and are generally referred to as alkalophilic _Bacillus_ strains. For the present

invention we define alkalophilic bacilli as <u>Bacillus</u> strains that grow well under alkaline conditions pH 9-11 (Horikoshi, K. and T. Akiba, 1982, Alkalophilic microorganisms, Springer Verlag, New York). The alkaline 5 proteases produced by such bacilli are also called high alkaline proteases. Examples of <u>Bacillus</u> strains capable of growing at alkaline pH are described in, for example, U.S. Patent Nos. 3,723,250, Re. 30,602 and 4,480,037. An example of an alkalophilic <u>Bacillus</u> host strain is <u>Bacillus</u> novo 10 species PB92 disclosed <u>inter</u> <u>alia</u> in U.S. Patent No. Re. 30,602. Derivatives of these alkalophilic <u>Bacillus</u> strains, that have been optimized for protease production are employed to produce their proteases on industrial scale (See EP-A-0284126). The products are used in several industrial 15 applications e.g. as an additive in laundry detergents. Examples of such products are Maxacal[R] (Gist-brocades/IBIS), Savinase[R] (NOVO), Esperase[R] (NOVO). Mutants of such products have been described in WO 89/06279, where it is said that they are derived from <u>Bacillus</u> <u>lentus</u> strains, and in the 20 not prepublished European Patent Application EP-A-0328229.

According to a preferred embodiment of the present invention transformed <u>Bacillus</u> strains are provided which produce the mutants described in EP-A-0328229 and WO 89/06279, preferably on an industrial scale.

25 According to the present invention a <u>Bacillus</u> strain is suitably used, in particular an alkalophilic <u>Bacillus</u>, preferably PB92. Another preferred group of <u>Bacillus</u> strains are the asporogenous mutants, of which PBT110 and its derivatives are most preferred. Transformation of 30 alkalophilic <u>Bacillus</u> strains will preferably involve the use of protoplasts from said strains. However, the conventional protoplast transformation protocol as described by Chang, S. and S.M. Cohen, Molec. Gen. Genet. <u>168</u> (1979) 111-115, does not work for alkalophilic <u>Bacillus</u> strains. 35 The protocols needed for these strains have been disclosed in EP-A-0283075, which is herein included by reference.

Expression of mutant protease genes in homologous hosts necessitates the replacement and/or inactivation of the wild type gene. Some methods can be used.

A) One method is to clone the gene or part thereof,
5 modify it by site-directed mutagenesis and reintroduce the (partial) gene into the cell on a plasmid. By homologous recombination the gene may be introduced into the chromosome. Resulting in a situation in which a wild-type and the mutant gene are tandemly located. After a second
10 recombination the modified sequence is left in the chromosome having thereby effectively introduced the mutation into the chromosomal gene (as described in Figure 14).

B) In another method the chromosomal gene copy is
15 inactivated by insertion or deletion. If insertional inactivation is chosen as a strategy the DNA fragment used as an insert may be any fragment as long as the formation of a complete protein product from the inactivated gene is prevented. Preferably the fragment contains a selectable
20 marker. If deletion of the gene is chosen as a strategy the chromosomal gene fragment that is deleted can be the whole coding region or a part thereof. After the inactivation a mutant copy of the inactivated gene is brought into the cell on a cloning vector.

25 C) In another method the chronosomal gene copy is mutagenized. In principle it is possible to obtain (specific) mutations in vivo, after transformation of bacteria with oligonucleotides which are mutagenic. Such a method has been succesfully used in Yeast as described by
30 Moerschell et al. Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 524-528. If such a method is performed on a homologous preferably a homologous alkalophilic bacillus strains this is embodiment of this invention.

The present invention relates to Bacillus strains
35 that are known to be efficient protease producers. In one of the preferred methods the complete alkaline protease coding region is deleted from the vector in vitro, leaving behind

the 5' and 3' flanking sequences. The vector used is preferably a plasmid. To make selection possible a selectable marker is introduced into the plasmid preferably an antibiotic resistance e.g. neomycin resistance.

5 Preferably the vector can selectively be integrated into the chromosome this can be achieved by introducing an inducible origin of replication e.g. a temperature sensitive origin like the one from pE194. Said plasmid is introduced into a homologous host cell. Under high temperature condition, the

10 plasmid is not able to replicate so that chromosomal integrants can be selected for. Integrants are selected at high temperatures in the presence of the relevant antibiotic (e.g. neomycin). Resolution of the plasmid from the host chromosome can leave the flanking regions in the

15 chromosome while removing the coding region. Finally, the resulting mutant selected for has lost the selectable marker gene (is neomycin sensitive) and is protease negative. The final chromosomal configuration is determined by restriction analysis and Southern blotting. WO 88/06623 extensively

20 describes the possible integration mechanisms.

The protease gene, modified by site-directed mutagenesis is now introduced into the protease negative cell on a suitable expression vector. Care is taken that the plasmid carries short or preferably no sequences homologous

25 to the chromosome to avoid plasmid instability.

In another preferred embodiment the flanking regions are left in the plasmid so that the whole of the mutated gene can become integrated in the chromosome. This is especially advantageous if the plasmid turns out be

30 unstable. The present invention also shows that the amount of recombinant mutant or wild-type protease obtained is comparable or even higher if one copy of the gene is integrated in the genome then if there are several copies of the gene encoded by a plasmid.

35 In yet another embodiment the chromosomal high alkaline protease gene is replaced by the mutant gene by

homologous recombination, without prior isolation of a protease negative strain.

The protease is expressed and can be isolated either from the cells or from the medium when the protein is
5 secreted.

Plasmid pM58, constructed as described in EP-A-0284126, contains the alkaline protease gene. To obtain the inactivation plasmid the coding region of the protease gene is deleted by BalI/HpaI double digestion. The temperature-
10 sensitive origin of replication from pE194neo-1 is subsequently introduced. The construct pEM58 is introduced into <u>Bacillus</u> strain PBT110.

Integration by the so-called Campbell-type mechanism took place in the 5' flanking region of the protease gene,
15 resulting in a protease positive strain carrying the entire plasmid vector in its chromosome in the protease locus (as depicted in Figure 3).

Subsequent selection for protease negative derivatives of this strain revealed two strains PBT125 and
20 PBT126, which had lost the protease gene and the neomycin marker. Further analysis showed that in this second step illegitimate recombination had occured.

Since the obtained <u>Bacillus</u> strains did not show a detectable protease activity they could be used for
25 expression of altered protease genes. It will be clear that PBT125 and PBT126 are only examples. Since illegitemite recombination will give different results it will be essential to determine whether the complete gene has been deleted. To obtain more specific strains, homologous
30 recombinants could have been made.

The examples given here mainly describe two specific embodiments of the invention. One is the production of plasmid encoded mutant protease from a protease negative strain. The other is the production of mutant protease
35 whereby the mutated gene has been introduced in the genome by gene exchange. It will immediately be clear to someone skilled in the art that other embodiments are possible, for

example; chromosomal integration in a protease negative strain, a mutated gene copy both on a plasmid and integrated into the chromosome, more than one copy of the mutated gene tandamly located or at different positions in the

5 chromosome.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL SECTION

EXAMPLE 1

5   Construction of Inactivation Plasmid pEM58 Δ

Plasmid pM58 (see EP-A-0284126) was digested with restriction enzymes BalI and HpaI. The large fragment containing part of the flanking sequences of the protease
10 gene was purified (Maniatis, Molecular cloning: A Laboratory Manual, Cold Spring Harbor 1982) and ligated. The ligation mix was transformed (Spizizen et al., J. Bacteriol. 81 (1961) 741-746) to Bacillus subtilis DB104 (Kawamura and Doi, J. Bacteriol. 160 (1984) 442-444) and selected for
15 neomycin resistant and protease negative transformants on minimal plates containing 0,4% casein and 20 µg/ml neomycin.

Plasmid pM58 Δ was isolated and characterized by restriction enzyme analysis. As shown in Figure 1 plasmid pM58 Δ contains the flanking sequences of the alkaline
20 protease gene, the gene coding for neomycin resistance and a Bacillus origin of replication.

To construct an integration plasmid containing a temperature sensitive origin of replication, pM58 Δ was digested with restriction enzymes XbaI and BglII. The
25 fragment containing the alkaline protease flanking sequences was purified and ligated to the (purified) XbaI/BglII fragment of pE194neo-1 (EP-A-0284126) containing the temperature sensitive origin of replication, derived from pE194 (Jordanescu et al., Plasmid 1 (1978) 468-479) after
30 said fragment was purified as formerly described. The ligation mix was transformed to B. subtilis DB104 and selected for neomycin resistance (see above). Plasmid pEM58 Δ was isolated and characterized (Figure 2). It contains the neomycin resistance gene, the temperature sensitive origin
35 of replication from pE194 and the flanking sequences of the alkaline protease gene.

This plasmid pEM58 $\triangle$ was used to transform <u>Bacillus</u> strain PBT110.

<center>EXAMPLE 2</center>

5

<center><u>Construction of a Protease Negative
Alkalophilic Bacillus Strain</u></center>

A. <u>Protoplast transformation of Bacillus PBT110 by plasmid</u>
10   <u>pEM58 $\triangle$</u>

    <u>Bacillus</u> strain PBT110 is an asporogenous mutant of <u>Bacillus</u> novo species PB92 (U.S. Patent No. Re. 30,602) and was obtained by classical (UV) mutation procedures. This
15 strain was transformed with plasmid pEM58 $\triangle$ similar to the method described in EP-A-0284126. Prior to integration experiments, transformants were checked by restriction enzyme analysis whether they contained the relevant plasmid.

20 B. <u>Integration of pEM58 $\triangle$ in Bacillus PBT110 chromosome</u>

    Integration of pEM58 $\triangle$ in <u>Bacillus</u> strain PBT110 chromosome was performed as described in EP-A-0284126. Selection for integrants occurred at a neomycin concen-
25 tration of 1 μg/ml. The genetic organization of the integrant was determined by restriction enzyme analysis followed by Southern blotting and hybridization analysis and is shown in Figure 3. It appeared that integration of pEM58 $\triangle$ took place through a so-called Campbell-type mechanism by
30 homologous recombination in the 5' flanking region of the alkaline protease gene, resulting in <u>Bacillus</u> strain PBT110-INT5.

C. <u>Selection for protease negative strains</u>

35

    <u>Bacillus</u> strain PBT110-INT5 was inoculated in 100 ml

tryptic soya broth (TSB) containing 1 μg/ml neomycin and incubated for 24 hours at 50°C.

After 24 hours 0.1 ml of the culture obtained was inoculated in a 500 ml shake flask containing 100 ml PBT minimal medium: K_2HPO_4, 17.42 g/l; glutamate, 13.36 g/l; citrate.H_2O 2 g/l; $FeSO_4.7H_2O$, 0.05 g/l; $ZnSO_4.7H_2O$, 1 mg/l; $MnSO_4.H_2O$, 1 mg/l; $CaCl_2.2H_2O$, 10 mg/l; $MgSO_4.7H_2O$, 0.2 g/l; $CuSO_4.5H_2O$, 0.5 mg/l; H_3BO_3, 0.5 mg/l; $Na_2MoO_4.2H_2O$, 0.5 mg/l; $CoCl_2.6H_2O$, 0.5 mg/l; Biotine, 0.5 mg/l; saccharose, 20 g/l; (pH 8.0, sterilized for 20 min at 120°C). After sterilization 1 mg/l thiamine was added.

This culture was incubated for 4 days at 37°C. After 4 days, 1 ml of culture was diluted in 100 ml of the same medium and incubated for 4 days. After 4 days the culture was plated on PBT minimal medium plates containing additionally 0.4% casein and 15 g/l agar. Colonies which lacked a detectable protease halo were considered protease minus and were checked on the absence of the gene encoding the PBT110 alkaline protease.

D. **Characterization of protease negative strains**

Strains which did not show a detectable halo on the casein plates described in the previous section were primarily tested on the presence of a neomycin sensitive and asporogenous phenotype. Two strains, Bacillus PBT125 and PBT126, which contained the protease negative and asporogenous phenotype were tested for protease production in 500 ml shake flasks containing 100 ml protease production medium, as described in U.S. Patent No. Re. 30,602. Neither strain produced detectable amounts of protease.

The genetic organization of the two strains was determined by restriction enzyme analysis and chromosomal blotting and is shown in Figure 4. No homologous but illegitimate recombination had occurred, resulting in two strains with completely deleted protease and neomycin resistance genes. However, still a small plasmid fragment of

undefined size appeared to be left in the chromosome after deletion of the protease gene. The sequence of this fragment was determined as follows. The chromosomal HindIII fragments of strains PBT125 and PBT126 containing the plasmid/chromosome junctions were ligated in phage vector M13 mp18 (Messing et al., Nucl. Acids Res. 9 (1981) 303-321) and transfected to E. coli JM101 according to the procedure described by Cohen et al., Proc. Natl. Acad. Sci. USA 69 (1972) 2110-2114. After phage propagation in E. coli JM101, ssDNA was isolated (Heidecker et al., Gene 10 (1980) 69-73).

The inserts were sequenced using the method described by Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977) 6463. The results are shown in Figures 5A and 5B, respectively.

EXAMPLE 3

Construction of Protease Production and Mutation Vectors

A. Subcloning of the PB92 protease gene in phage M13 mp18

Plasmid pM58 was digested with HpaI and BalI. After purification of the DNA fragment containing the protease gene (Maniatis, 1982) this fragment was ligated into phage M13 mp18 which was digested with SmaI. The ligation mixture was transfected to E. coli JM101 (Cohen et al., supra). After phage propagation in E. coli JM101, dsDNA was isolated according to the method described by Birnboim and Doly (Nucl. Acids Res. 7 (1979) 1513-1523). The insert and its orientation were checked by restriction enzyme analysis. Vector mp18-MXL was used for further subcloning experiments and is shown in Figure 6.

B. Subcloning of the PB92 protease gene in plasmid pBHA1

Figure 7 shows the nucleotide sequence of plasmid pBHA1. This plasmid is derived from the twin vector system

pMa/c5-8, described by Stanssens <u>et</u> <u>al</u>. (Nucleic Acids Res. <u>17</u> (1989) 4441-4454) in which an additional promoter was inserted. Plasmid BHA1 consists of the following fragments:

5   pos 11-105:   bacteriophage FD, terminator;

pos 121-215:   bacteriophage FD, terminator;

pos 221-307:   a part of plasmid pBR322 (viz. positions 2069-2153);

pos 313-768:   bacteriophage F1, origin of replication (viz.
10   pos 5482-5943);

pos 772-2571:   part of plasmid pBR322, viz. the origin of replication and the β-lactamase gene;

pos 2572-2685: transposon Tn903, complete genome;

pos 2719-2772: tryptophan terminator (double);

15   pos 2773-3729: transposon Tn9, the chloramphenicol-acetyl-transferase (=cat) gene. The nucleotides at pos 3005(A), 3038(C), 3302(A) and 3409(A) differ from the wild-type cat coding sequence. The mutations were introduced to
20   eliminate the <u>NcoI</u>, <u>BalI</u>, <u>EcoRI</u> and <u>PvuII</u> sites;

pos 3730-3804: multiple cloning site;

pos 3807-7264: part of plasmid pUB110 (viz. the replication function and kanamycin resistance gene,
25   <u>EcoRI</u>-<u>PvuII</u> fragment) (McKenzie <u>et</u> <u>al</u>., Plasmid <u>15</u> (1986) 93-103 and Plasmid <u>17</u> (1987) 83-85);

pos 7267-7331: multi cloning site.

30   The fragments were combined using known cloning techniques, e.g., filling in sticky ends with Klenow, adapter cloning, etc. All data were derived from Genbank[R] National Nucleic Acid Sequence Data Bank' NIH, U.S.A. Plasmid pMc5-8 was deposited under DSM 4566.

35

     Prior to mutation procedures the twin vector system

pBHA/C-1 was modified to obtain the twin vector system pBHARB-MXL/pBHCRB-MXL, as follows:

1. Vector mp18-MXL was digested with KpnI and HincII. The DNA fragment containing the protease gene was purified and ligated into pBHA1 which was digested with EcoRV and KpnI. The ligation mixture was transformed to E. coli JM101 (Maniatis, 1982) and plated on LC$^+$ plates containing 10 g/l trypton, 5 g/l Yeast extract (Difco), 8 g/l NaCl, 25 g/l thymine, 1 g/l MgSO$_4$·7H$_2$O, 100 μg/ml ampicillin and 20 μg/ml neomycin pH 7.0. Plasmid DNA of transformants was isolated (Birnboim, supra) and characterized by restriction enzyme analysis. In this way pBHA1-MXL was isolated (see Figure 8).

2. In order to sequence mutations introduced in the protease gene with an available set of oligonucleotides, it was necessary to reverse the orientation of the F1 origin compared to the protease gene. This was performed in the following way: plasmid pBHA1-MXL was digested with BamHI and religated. The ligation mix was transformed to E. coli WK6 (Maniatis, supra) and selected for neomycin or ampicillin resistance on LC$^+$ plates containing 10 g/l trypton, 5 g/l Yeast Extract (Difco), 8 g/l NaCl, 25 mg/l thymine, 1 g/l MgSO$_4$.7H$_2$O, 100 μg/ml ampicillin and 20 μg/ml neomycin pH 7.0. Plasmid DNA of transformants was isolated (Birnboim, supra) and characterized by restriction enzyme analysis. In this way pBHAR-MXL was isolated, which differs from pBHA-MXL in that the orientation of the E. coli sequence is reversed, (see Figure 9).

3. For the introduction of additional flanking sequences in plasmid pBHAR-MXL, the following digestions were made. Plasmid pM58 was digested with SphI and HindIII. Plasmid pBHAR-MXL (Figure 10) was digested with SphI and HindIII. The larger fragment was purified. Both digests were ligated and transformed to B. subtilis DB104 and selected for neomycin resistance and protease activity on minimal plates containing 10 μg/ml neomycin and 0.4 % casein. Transformants were characterized by restriction enzyme

analysis. One of these, pBHARB-MXL (Figure 10), was used for further experiments.

C. Mutagenesis of the PB92 gene in plasmid pBHARB-MXL
5

Mutagenesis was carried out with the twin vector system pBHARB-MXL/pBHCRB-MXL (Stanssens, supra).

A method based on the gapped-duplex approach (Kramer et al., Nucl. Acids Res. 12 (1984) 9441) and a phasmid
10 (phage/plasmid hybrid) was used. Essentially the method rests on a gapped-duplex DNA intermediate consisting of a gapped strand (- strand) containing a wild-type antibiotic resistance marker and a template strand (+ strand) carrying an amber mutation in the gene conferring resistance to the
15 antibiotic. After annealing, the mutagenic oligonucleotide becomes incorporated in the gapped strand during in vitro gap-filling and sealing reaction. The resultant molecules are used to transform a mismatch repair deficient (Mut S) host in which the linkage between the intended mutation and
20 the antibiotic resistance marker is preserved. The mixed phasmid population, isolated from this strain, is then allowed to segregate in a suppressor negative host strain. Transformants are plated on antibiotic containing medium, thus imposing a selection for progeny derived from the
25 gapped strand.

In the pMa type vector nucleotide 3409 is changed from G to A, while in the pMc type vector nucleotide 2238 is changed from G to C, creating amber stopcodons in the chloramphenicol-acetyltransferase gene and β-lactamase
30 gene, respectively, rendering said genes inactive.

To perform mutagenesis the target DNA fragment is cloned into the multiple cloning site of pMa5-8 or a derivative thereof. A gapped duplex between pMa5-8
35 containing the target DNA and pMc5-8 is then constructed.

The single strand gap, consisting of the target DNA, may be subjected to mutagenesis with a mutagenic oligo-

nucleotide, with long synthetic oligonucleotides having a low level of misincorporated nucleotides, using chemical or enzymatic misincorporation of nucleotides. For a detailed description, see Ausubel et al., 1987, Current Protocols in
5 Molecular Biology, John Wiley & Sons Inc. New York; or B. Perbal, 1988, A practical Guide to Molecular Cloning, 2nd ed. John Wiley & Sons Inc; New York.

D.   Construction of Vectors for Industrial Protease
10   Production

    After the introduction of the desired mutation(s) in plasmid pBHARB-MXL, the unwanted E. coli sequences from the plasmid were deleted as follows: plasmid pBHARB-MXL
15 containing the relevant mutation was digested with BamHI and religated under diluted conditions. The ligation mixture was transformed to B. subtilis DB104 and selected for neomycin resistance and protease activity on minimal plates containing 20 μg/ml neomycin and 0.4 % casein. DNA of
20 transformants was isolated and characterized by restriction enzyme analysis. In this way vectors were isolated lacking E. coli DNA, which are suitable for commercial production of proteins.
    The foregoing procedure is illustrated in Figure 11
25 where mutation M216Q is taken as an example, resulting in plasmid pBHB-MXL M216Q. M216Q, and also M216S, S160D and N212D referred to hereinafter, are mutant proteases of Bacillus PB92, described in EP-A-0328229.

30   EXAMPLE 4

                Protoplast Transformation of Bacillus PBT125
                    by Plasmid pBHB-MXL Derivatives

35   Bacillus strain PBT125 was transformed with plasmid pBHB-MXL or plasmids derived thereof containing mutant protease genes, following the transformation procedure

described in EP-A-0284126. Prior to production studies, transformants were checked by restriction enzyme analysis to contain the relevant plasmid.

5 <div align="center">EXAMPLE 5</div>

<div align="center">Construction of Protease Integration Vectors</div>

A.   Integration vector pEN₃Q

10

Plasmid pE194neo-1 (EP-A-0284126) was digested with SalI and religated. The ligate was transformed to B. subtilis DB104 and selected for neomycin resistance on minimal plates containing 20 µg/ml neomycin. Plasmid pE194

15 neo-3 was isolated and checked on the presence of the neomycin resistance gene in a reversed orientation (see Figure 12). Plasmid pE194neo-3 was digested with HpaI and BglII. The fragment containing the temperature sensitive origin of replication was purified (Maniatis, supra).

20 Plasmid pBHB-MXL M216Q (see Example 3 and Figure 11) was digested with BglII and BalI. The fragment containing the protease gene was purified. Both fragments were ligated. The ligation mix was transformed to B. subtilis DB104 and selected for protease activity and neomycin resistance as

25 described before. Transformants were characterized by restriction enzyme digestions. One of these, containing plasmid pEN₃Q, was selected (Figure 13). This plasmid contained the neomycin resistance gene, the t.s. origin of replication from plasmid pE194neo-3 and the M216Q mutation

30 in the PB92 protease gene.

B. Integration vector pEN₃S.

In the same way as described for integration vector

35 pEN₃Q, pEN₃S was constructed using pBHARB-MXL M216S as starting vector.

EXAMPLE 6

Construction of Bacillus Strains PEP111 and PEP112

5 A. Protoplast transformation of Bacillus PBT110 by Plasmid pEN$_3$Q

Bacillus strain PBT110 was transformed with plasmid pEN$_3$Q as described in EP-A-0284126. Prior to integration
10 experiments, it was checked by restriction enzyme analysis whether the transformants contained the relevant plasmid.

B. Integration of pEN$_3$Q in Bacillus PBT110 chromosome

15 Integration experiments with plasmid pEN$_3$Q in Bacillus strain PBT110 chromosome were performed as described in EP-A-0284126. Selection for integrants occurred at a neomycin concentration of 1µg/ml at non-permissive temperatures (50°C) for plasmid replication. To
20 check the integration of pEN$_3$Q in the chromosome, chromosomal DNA of potential integrants was isolated, digested with HindIII or ClaI, run on a 0.8% DNA agarose gel and blotted to nitrocellulose (Southern, J. Mol. Biol. 98 (1975) 503-517) and hybridized with ^{32}P labeled nick-
25 translated pEN$_3$Q DNA (Maniatis, 1982).

It was shown that integration of pEN$_3$Q had occurred by homologous recombination resulting in a strain (PBT110M/Q) with two protease genes (one wild type and one mutant M216Q) tandemly located on the chromosome. Thus,
30 integration of pEN$_3$Q occurred by a so-called Campbell-type mechanism as depicted in Figure 14 (see also EP-A-0284126).

C. Selection of neomycin sensitive recombinants

35 Neomycin sensitive recombinants were selected, which contain either the wild-type or the mutant protease gene in the chromosome (hereinafter called "outrecombinants"). The

selection procedure is similar to the procedure described before in Example 2C.

Strain PBT110 M/Q was used as starting strain. After strain PBT 110 M/Q was incubated in PBT minimal medium, the
5 culture was plated on PBT minimal medium plates containing 0.4% casein and 15 g/l agar. These plates were incubated for 48 h at 37°C and replica plated to Heart infusion (HI) plates containing 1 μg/ml neomycin and no neomycin, respectively. Neomycin sensitive colonies were considered
10 outrecombinants.

D. Characterization of neomycin sensitive recombinants.

Chromosomal DNA of potential outrecombinants was
15 isolated, digested with ClaI or HindIII, run on a 0.5% agarose gel, blotted to nitrocellulose (Southern, 1975) and hybridized with ^{32}P labeled nick translated pEN$_3$Q (Maniatis, 1982). Since mutation M216Q results in the removal of a ClaI site, the strains containing a wild-type
20 protease or a mutant protease gene, and the intermediate strain PBT110 M/Q can be easily distinguished, see Fig. 14.

An outrecombinant containing the M216Q mutant protease gene of PB92 was isolated. After it was shown that a gene replacement of PB92 wild-type protease to PB92 mutant
25 protease had occurred, the product of the strain was characterized by its biochemical parameters (k_{cat}, K_m), oxidation resistance, specific activity and wash performance, as described EP-A-0328229. All results were equal to a control protease M216Q, indicating that the
30 product of the transformed strain was indeed the PB$_{92}$ protease containing the mutation M216Q. This transformed strain is called Bacillus PEP111.

Similar experiments were performed using vector pEN$_3$S to construct Bacillus strain PEP112, which contains
35 the PB92 protease with mutation M216S.

EXAMPLE 7

Construction of Bacillus Strains PEP211 and PEP212

5   Bacillus strain PEP111 was transformed with plasmid pEN$_3$Q, following the procedure described in EP-A-0284126. Prior to the integration procedure, it was checked by restriction enzyme analysis whether the transformants contained the relevant plasmid. The integration procedure 10 and the selection for integrants (at a neomycin concentration of 20 μg/ml at non-permissive temperatures (50°C) for plasmid replication) was performed as described in EP-A-0284126.

To check the integration of plasmid pEN$_3$Q in the 15 chromosome, chromosomal DNA of potential integrants was isolated and digested with HindIII, run on 0.8% agarose gels, blotted to nitrocellulose (Southern et al. 1979) and hybridized with ^{32}P labeled nick translated pEN$_3$Q. It was shown that a strain containing two mutant (M216Q) PB$_{92}$ 20 protease genes, separately located on the chromosome was isolated. This strain is called PEP211. Its chromosomal organization is shown in Figure 15.

To obtain an analogous strain containing two protease genes with mutation M216S, strain PEP112 containing a mutant 25 gene (M216S) after gene replacement of the wild-type gene, was transformed with plasmid pEN$_3$S and procedures were followed as described above. These experiments resulted in strain PEP212 containing two separately located PB$_{92}$ mutant M216S protease genes on the chromosome.
30

EXAMPLE 8

Production of Mutant Proteases

35   Production of protease mutants was performed with transformed Bacillus strains in which the mutated protease

encoding gene was either localized on a plasmid or on the chromosome.

For plasmid encoded production 2 types of strains were used, the alkalophilic strain Bacillus PBT125, as

5   herein described, and B. subtilis strain DS12367, an asporogenic strain derived from B. subtilis DB104 (Kawamura et al., J. Bacteriol. (1984) 160, 442-444. These strains were transformed with pBHB-MXL derived plasmids. Bacillus PBT125 and B. subtilis DS12367, which do not possess

10 protease encoding plasmids, were used as control strains.

For production with strains containing chromosomally integrated protease mutant genes the Bacillus PEP strains, as herein described, were used. Bacillus PBT strains 108 and 110, containing the wild-type protease genes were used as

15 control strains. PBT108 is a strain containing two wild-type genes separately located on the chromosome, see EP-A-0284126. For PBT110, see Example 2A.

To check the activity of the original promoter of the PB92 protease gene in B. subtilis, a construct was made,

20 pBHB-MXLR, in which the orientation of the protease gene to the HpaII promoter (Zyprian et al., DNA 5 (1986) 219-225) was reversed, so that the expression of the protease gene is only directed by its original promoter.

The productions were performed in 500 ml shake flasks

25 containing 100 ml protease-production medium.

When alkalophilic Bacillus strains were used cultures were incubated, after inoculation, for 40 hours at 37°C in a production medium as described in U.S. Patent No. Re. 30,602. In case of plasmid containing strains, neomycin (20

30 μg/ml) was added.

When B. subtilis DS12367 strains were used, a comparable medium containing 12.5 g/l Yeast Extract (Difco), 0.97 g/l $CaCl_2.6H_2O$, 2.25 gl $MgCl_2.6H_2O$, 20 mg/l $MnSO_4.4H_2O$, 1 mg/l $CoCl_2.6H_2O$, 0.5 g/l citrate, 0.5 ml antifoam 5693, 6%

35 w/w maltose, 0.2 M phosphate buffer pH 6.8 was applied. In case of plasmid containing strains, neomycin (20 μg/ml) was added.

Cultures containing <u>B</u>. <u>subtilis</u> strains were incubated for 65 hours at 37°C. In all cases the shake-flasks were inoculated with 0.1 ml of a Tryptic Soya Broth culture containing 20 μg/ml neomycin containing the relevant

5 strain which had been incubated for 24 hours at 37°C.

Protease activity was assayed using dimethylcasein as substrate as described by Lin <u>et</u> <u>al</u>., J. Biol. Chem. <u>244</u> (1969) 789-793. The specific activities of the protease mutants (EP-A-0328229) were used to determine the production

10 on mg basis.

The results of the fermentation experiments are summarized in the following Table 1.

<u>Table 1</u>

15

<u>Strain</u>	<u>Relative protease production</u> (%)	<u>Mutation</u>
PBT110	100	WT
PBT125	0	--
PBT108	120	WT
PEP111	100	M216Q
PEP112	100	M216S
PBT125 pBHB-MXL	95-100	WT
PBT125 pBHB-MXL M216Q	95-100	M216Q
PBT125 pBHB-MXL M216S	95-100	M216S
PBT125 pBHB-MXL S160D	95-100	S160D
PBT125 pBHB-MXL N212D	95-100	N212D
PEP211	120	M216Q
PEP212	120	M216S
DS12367	0-1	--
DS12367 pBHB-MXL	4	WT
DS12367 pBHB-MXL	40	WT
DS12367 pBHB-MXL M216Q	40	M216Q
DS12367 pBHB-MXL M216S	40	M216S
DS12367 pBHB-MXL S160D	40	S160D
DS12367 pBHB-MXL N212D	40	N212D

All publications (including patents and patent applications) mentioned in this specification are

45 indicative to the level of skill of those skilled in the art to which this invention pertains. All publications are

herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in
5 some detail by way of illustration and example for purposes of clarity of understanding, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

10

CLAIMS

1. A method for the production of proteases differing in at least one amino acid from the wild-type protease, said method comprising the use of a homologous alkalophilic Bacillus strain as expression host said strain being unable to produce the wild-type protease.

2. A method according to Claim 1, employing as expression host a protease negative alkalophilic Bacillus strain.

3. A method according to Claim 2, employing as expression host a protease negative alkalophilic Bacillus strain, wherein the wild-type protease gene is deleted from the genome.

4. A method according to any one of Claims 1 to 3, employing a protease negative derivative of Bacillus novo species PB92.

5. A method according to any one of Claims 1 to 4, employing an asporogenic alkalophilic Bacillus strain.

6. A method according to any one of Claims 1 to 5, employing an asporogenic Bacillus from which the wild-type protease gene has been deleted by homologous or illegitimate recombination.

7. A method according to any one of Claims 2 to 6, wherein the mutant high alkaline protease is plasmid encoded.

8. A method according to Claim 1, 2 or 3, wherein the wild-type high alkaline protease has essentially the amino acid sequence of Bacillus novo species PB92 protease.

9. A method according to Claim 1, wherein the wild-type alkaline protease gene is deleted from the chromosome, and at least one copy of a mutant high alkaline protease gene is inserted into the genome.

10. A method according to Claim 9, wherein at least one copy of a mutant high alkaline protease gene is inserted into the genome and another copy is plasmid encoded.

11. A method of obtaining an alkalophilic _Bacillus_ strain having a reduced extracellular alkaline protease level comprising:
   - isolating a gene, including the 5' and/or 3' noncoding regions thereof, coding for alkaline protease,
   - inserting the gene into a cloning vector,
   - deleting the coding region of the gene from the cloning vector,
   - transforming a _Bacillus_ strain with the cloning vector,
   - growing the transformants under conditions in which the replication function of the vector is inactive,
   - selecting transformants which have lost the DNA sequence encoding the alkaline protease.

12. A method according to Claim 11, wherein said alkalophilic _Bacillus_ strain is _Bacillus_ novo species PB92 or a derivative thereof.

13. An alkalophilic _Bacillus_ strain obtained according to Claim 11, transformed with a plasmid expression vector containing a mutant high alkaline protease gene.

14. A _Bacillus_ strain according to Claim 13, wherein the alkalophilic _Bacillus_ strain is a derivative of _Bacillus_ novo species PB92, obtained by homologous or illegitimate recombination.

15. An alkaline protease differing in at least one amino acid from the wild-type protease produced by a <u>Bacillus</u> strain having at least one copy of the mutated gene and no copy of the wild-type gene giving an active protease.

16. An alkaline protease according to Claim 15, produced by an alkalophilic <u>Bacillus</u> strain.

17. An alkaline protease according to Claim 15 or 16, where the mutant gene is provided on a plasmid and/or in the genome.

EFFICIENT PRODUCTION OF MUTANT PROTEASES

5 ## Abstract of the Disclosure

Alkalophilic <u>Bacillus</u> strains are provided which efficiently produce mutant <u>Bacillus</u> proteases and which are incapable of expressing the wild-type proteases.

11 06 89

Figur 1

Bacillus novo PB92 DNA

plasmid DNA

HpaI
5'
BalI
3'

protease gene

pM58
6.3kb

Neo

ori

x HpaI/BalI purification of the large
Fragment
Ligation

Transformation to B.subtilis DB104.
selection: prot$^\oplus$, Neo^R .

BalI/HpaI
5'
3'

pM58Δ
4.9kb

Neo

ori
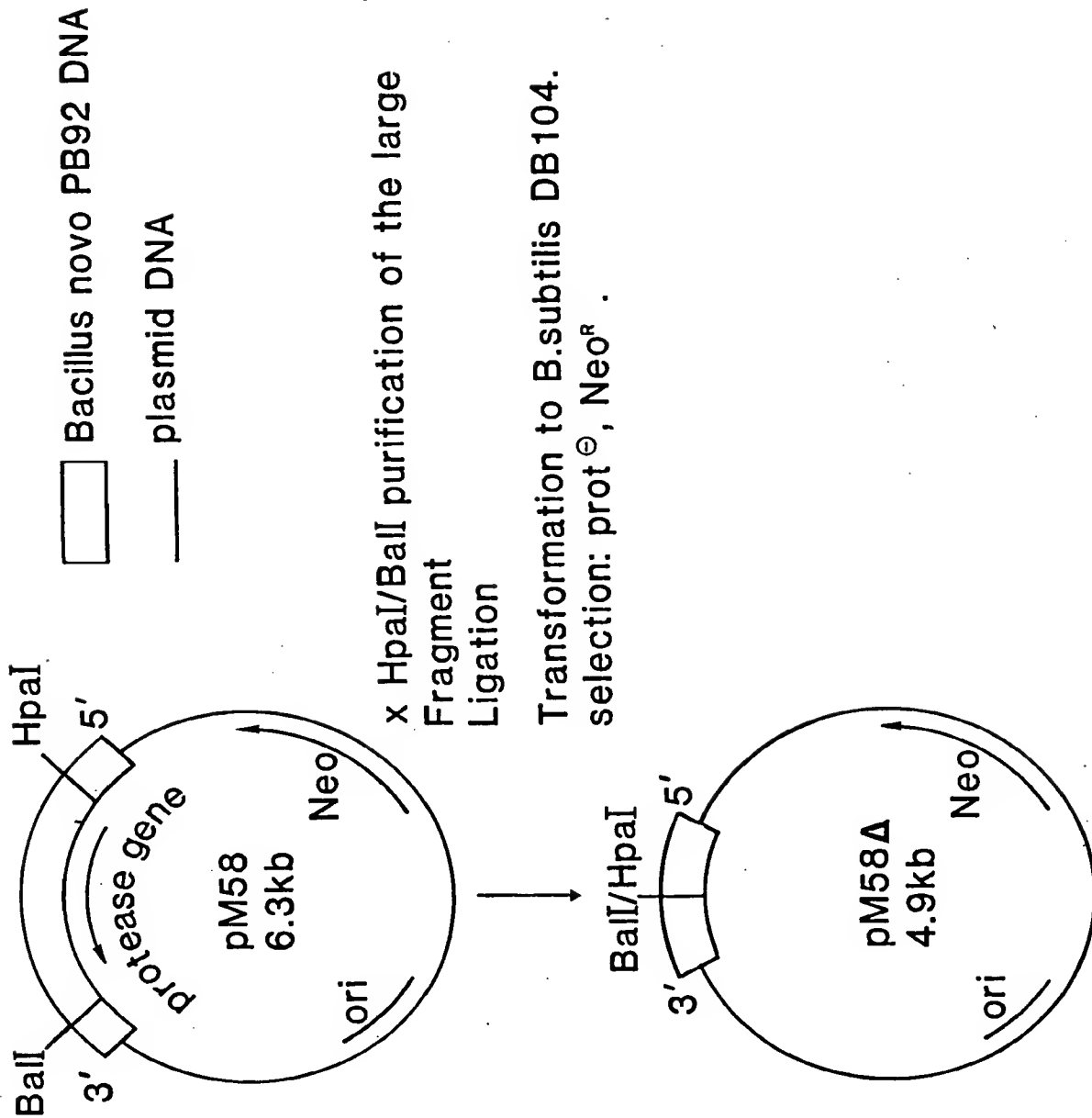
Figure 2

Figure 3

Figure 4.

```
HindIII                                40                              .   .         .80
AAGCTTTTCAAATGATCCAGGAGAACAAAAACGATTTTTTGAGGAAAGTTATAAATTATTTTCCGAACGATATGGCAAGC

                .             .            120                                      160
AAAATATTGCTTATGCAACAGTTCATAATGATGAGCAAACCCCTCACATGCATTTAGGTGTTGTGCCTATGCGTGATGGA

                .             .            200                                      240
AAACTGCAAGGAAAAAATGTGTTTAATCGTCAAGAACTGTTATGGCTACAAGATAAATTCCCCGAGCATATGAAAAAACA

                .             .            280                                      320
GGGTTTTGAGTTGAAGCGTGGTGAACGTGGCTCTGACCGTAAACATATTGAGACAGCTAAATTTAAAAAACAAACTTTGG

                  XbaI                     360                                      400
AAAAAGAGATTGATTTTCTAGAAAATGAATACGGTAAGGAAAATATGCTGTATGCGACTGTCCATCTGGATGAAAGAGTC

                .             .            440                                      480
CCACATATGCACTTTGGTTTTGTCCCTTTAACAGAGGACGGGAGATTGTCTGCAAAAGAACAGTTAGGCAACAAGAAAGA

                .             .            520       .                              560
CTTTACTCAATTACAAGATAGATTTAATGAGTATGTGAATGAGAAAGGTTATGAACTTGAAAGAGGCACGTCCAAAGAGG

                .             .            600             .                        640
TTACAGAACGAGAACATAAAGCGATGGATCAGTACAAGAAAGATACTGTATTTCATAAACAGGAACTGCAAGAAGTTAAG

                .             .            680             .                        720
GATGAGTTACAGAAGGCAAATAAGCAGTTACAGAGTGGAATAGAGCATATGAGGTCTACGAAACCCTTTGATTATGAAAA

                .             .            760             .                        800
TGAGCGTACAGGTTTGTTCTCTGGACGTGAAGAGACTGGTAGAAAGATATTAACTGCTGATGAATTTGAACGCCTGCAAG

                  PstI       .            840             .                        880
AAACAATCTCTTCTGCAGAACGGATTGTTGATGATTACGAAAATATTAAGAGCACAGACTATTACACAGAAAATCAAGAA

                .             .            920             .                        960
TTAAAAAAACGTAGAGAGAGTTTGAAAGAAGTAGTGAATACATGGAAAGAGGGGTATCACGAAAAAAGTAAAGAGGTTAA

                .             .            1000            .                        1040
TAAATTAAAGCGAGAGAATGATAGTTTGAATGAGCAGTTGAATGTATCAGAGAAATTTCAAGCTAGTACAGTGACTTTAT

                .             .            1080            .                        1120
ATCGTGCTGCGAGGGCGAATTTCCCTGGGTTTGAGAAAGGGTTTAATAGGCTTAAAGAGAAATTCTTTAATGATTCCAAA

                .             .            1160            .                        1200
TTTGAGCGTGTGGGACAGTTTATGGATGTTGTACAGGATAATGTCCAGAAGGTCGATAGAAAGCGTGAGAAACAGCGTAC

                .             .            1240            .                        1280
AGACGATTTAGAGATGTAGAGGTACTTTTATGCCGAGAAAACTTTTTGCGTGTGACAGTCCTTAAAAATATACTTAGAGCG

                .             .            1320            .                        1360
TAAGCGAAAGTAGTAGCGACAGCTATTAACTTTCGGTTGCAAAGCTCTAGGATTTTTAATGGACGCAGCGCATCACACGC

                .             .            1400            .                        1440
AAAAAGGAAATTGGAATAAATGCGAAATTTGAGATGTTAATTAAAGACCTTTTTGAGGTCTTTTTTTTCTTAGATTTTTGG
                         plasm.seq.  chrom.seq.
                            ─────┼─────
                .             .                 .                                   1520
GGTTATTTAGGGGAGAAAACATAGGGGGGTACTACGACCAAAATGGAGGAATACGGCAGCGCGAGCATGGTGGATGCTTG

                .             .            1560            .                        1600
AATTTCTAGGCAACAACAACGCCGCTATTCTTGACGGTGGGATTCAGGCTTGGAAGAACAGTGGCTACCCGACAACGGGA

        HindIII
TGAGGTGAAGCTT
```

figure 5A

```
HindIII          .              .              40           . ..  .:: .  ... .    .. .. ...  .80
AAGCTTTTCAAATGATCCAGGAGAACAAAAACGATTTTTTGAGGAAAGTTATAAATTATTTTCCGAACGATATGGAAGC

        .              .              120          .              .              160
AAAATATTGCTTATGCAACAGTTCATAATGATGAGCAAACCCCTCACATGCATTTAGGTGTTGTGCCTATGCGTGATGGA

        .              .              200          .              .              240
AAACTGCAAGGAAAAAATGTGTTTAATCGTCAAGAACTGTTATGGCTACAAGATAAATTCCCCGAGCATATGAAAAAACA

        .              .              280          .              .              320
GGGTTTTGAGTTGAAGCGTGGTGAACGTGGCTCTGACCGTAAACATATTGAGACAGCTAAATTTAAAAAACAAACTTTGG

        .         XbaI         .              360          .              .              400
AAAAAGAGATTGATTTTCTAGAAAATGAATACGGTAAGGAAAATATGCTGTATGCGACTGTCCATCTGGATGAAAGAGTC

        .              .              440          .              .              480
CCACATATGCACTTTGGTTTTGTCCCTTTAACAGAGGACGGGAGATTGTCTGCAAAAGAACAGTTAGGCAACAAGAAAGA

        .              .              520          .              .              560
CTTTACTCAATTACAAGATAGATTTAATGAGTATGTGAATGAGAAAGGTTATGAACTTGAAAGAGGCACGTCCAAAGAGG

        .              .              600          .              .              640
TTACAGAACGAGAACATAAAGCGATGGATCAGTACAAGAAAGATACTGTATTTCATAAACAGGAACTGCAAGAAGTTAAG

        .              .              680          .              .              720
GATGAGTTACAGAAGGCAAATAAGCAGTTACAGAGTGGAATAGAGCATATGAGGTCTACGAAACCCTTTGATTATGAAAA

        .              .              760          .              .              800
TGAGCGTACAGGTTTGTTCTCTGGACGTGAAGAGACTGGTAGAAAGATATTAACTGCTGATGAATTTGAACGCCTGCAAG

        .         PstI   .              840          .              .              880
AAACAATCTCTTCTGCAGAACGGATTGTTGATGATTACGAAAATATTAAGAGCACAGACTATTACACAGAAAATCAAGAA

        .              .              920          .              .              960
TTAAAAAAACGTAGAGAGAGTTTGAAAGAAGTAGTGAATACATGGAAAGAGGGGTATCACGAAAAAAGTAAAGAGGTTAA

        .              .              1000         .              .              1040
TAAATTAAAGCGAGAGAATGATAGTTTGAATGAGCAGTTGAATGTATCAGAGAAATTTCAAGCTAGTACAGTGACTTTAT

        .              .              1080         .              .              1120
ATCGTGCTGCGAGGGCGAATTTCCCTGGGTTTGAGAAAGGGTTTAATAGGCTTAAAGAGAAATTCTTTAATGATTCCAAA

        .              .              1160         .              .              1200
TTTGAGCGTGTGGGACAGTTTATGGATGTTGTACAGGATAATGTCCAGAAGGTCGATAGAAAGCGTGAGAAACAGCGTAC

        .              .              1240         .              .              1280
AGACGATTTAGAGATGTAGAGGTACTTTTATGCCGAGAAAACTTTTTGCGTGTGACAGTCCTTAAAATATACTTAGAGCG

        .              .              1320         .              .              1360
TAAGCGAAAGTAGTAGCGACAGCTATTAACTTTCGGTTGCAAAGCTCTAGGATTTTTAATGGACGCAGCGCATCACACGC

        .              .              1400         .              .              1440
AAAAAGGAAATTGGAATAAATGCGAAATTTGAGATGTTAATTAAAGACCTTTTTGAGGTCTTTTTTTTCTTAGATTTTTGG
                         plasm.seq. chrom.seq.
                              ———|———
        .              .              .              .              .              1520
GGTTATTTAGGGGAGAAAACATAGGGGGGTACTACGACCATTACGGAGAACGCAAAGCAGTAATCTCTGTAGTGTATTGA

        .              .              1560         .              .              1600
AGCAGG.....CA.100BP.....TTCAGTCCGTCAAGAGATGCGCAACTACGATAACTCCATTCTTTATGGTATGGCTG

        .              .              1640         .              .              1680
AGAAAATAGCGATTATGAACGTCGTTGGCGGTACATTTACCGACGAGCCATATGGGGATTGCTGTCCGCAAACGGCGACG

     EcoRI    .              .              1720    HindIII
ACGAATTCACGCGAAAAAAATCAATGAAATTCTTGCGAGACTTAAAAGCTT                figure 5B
```
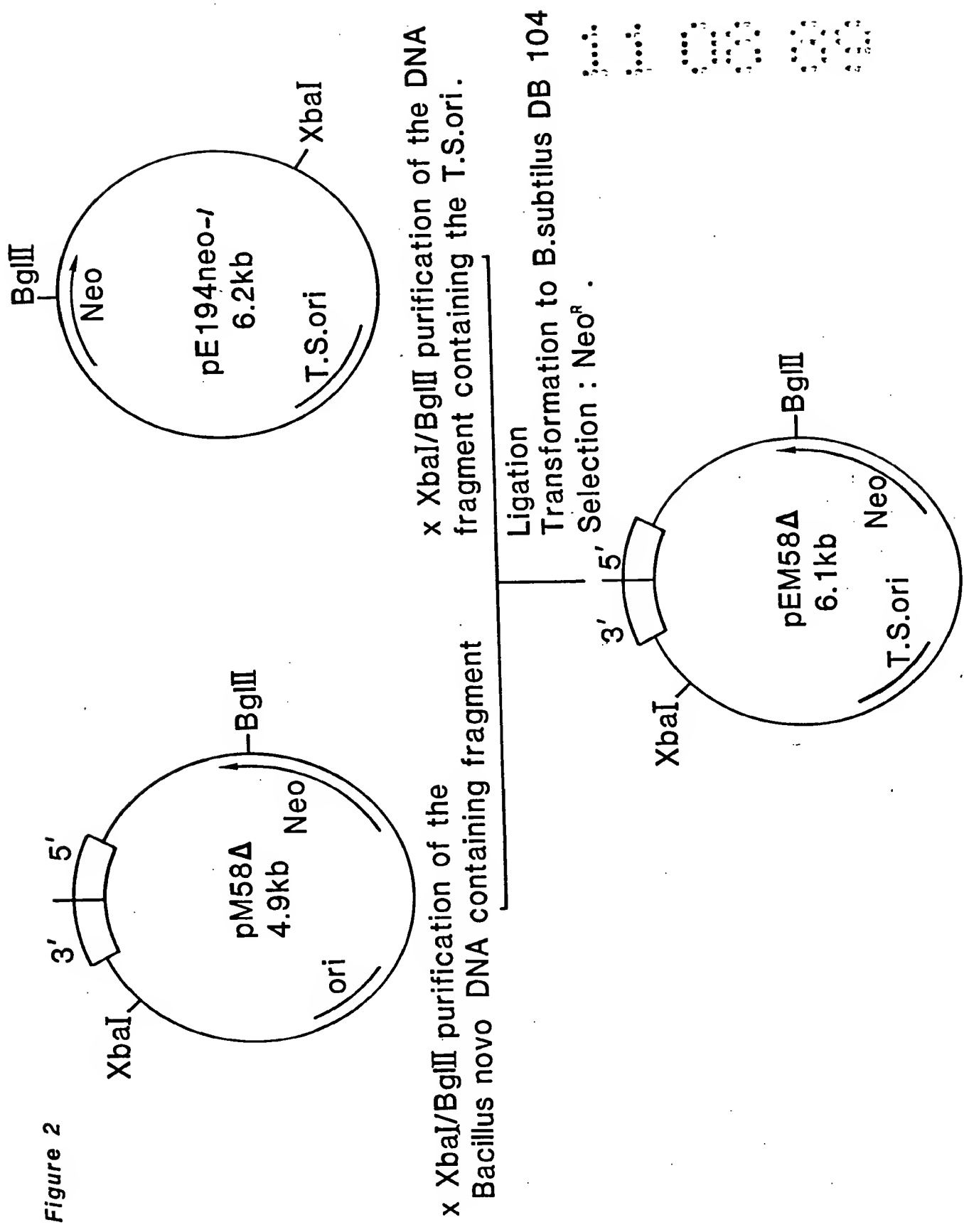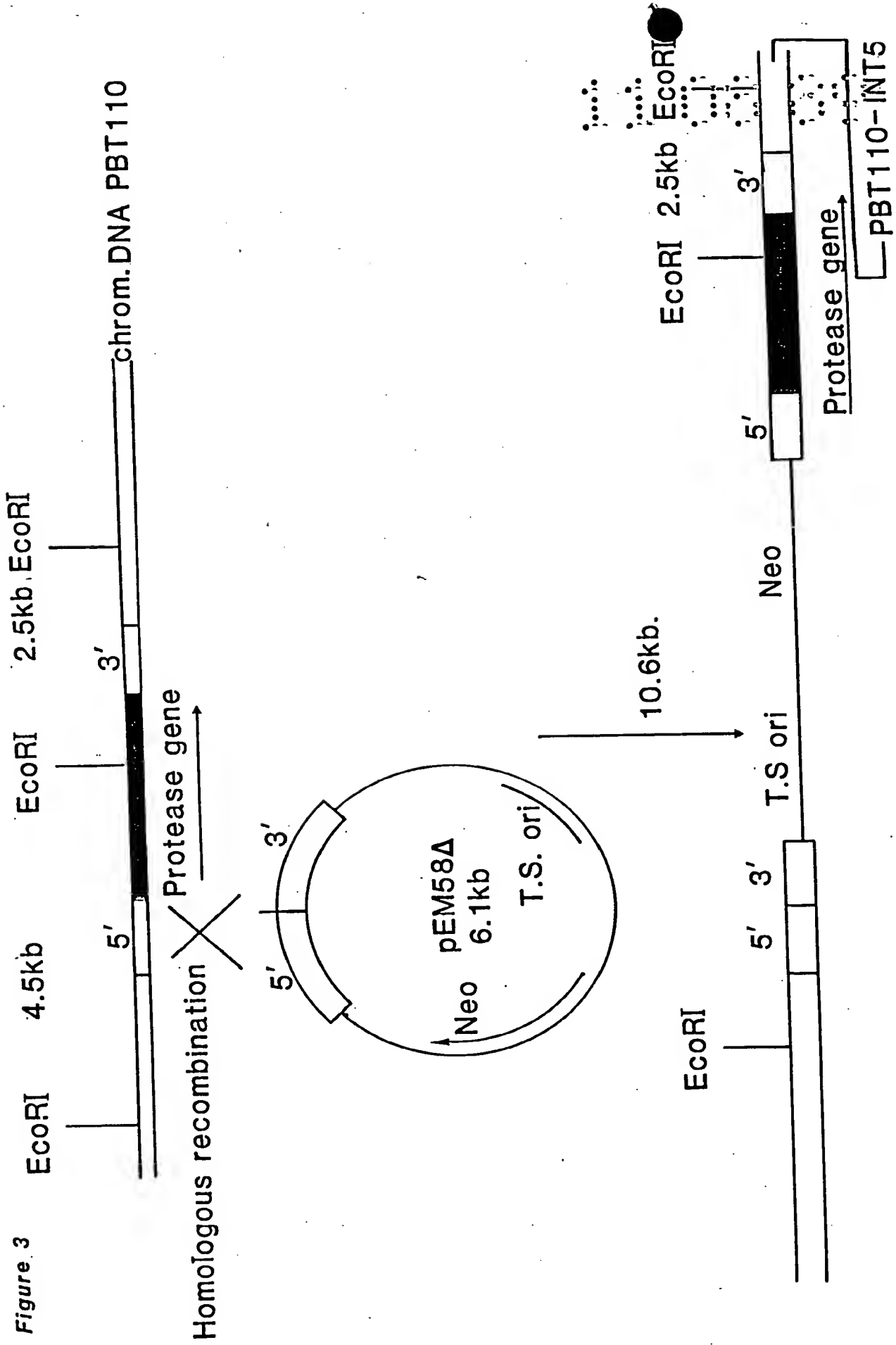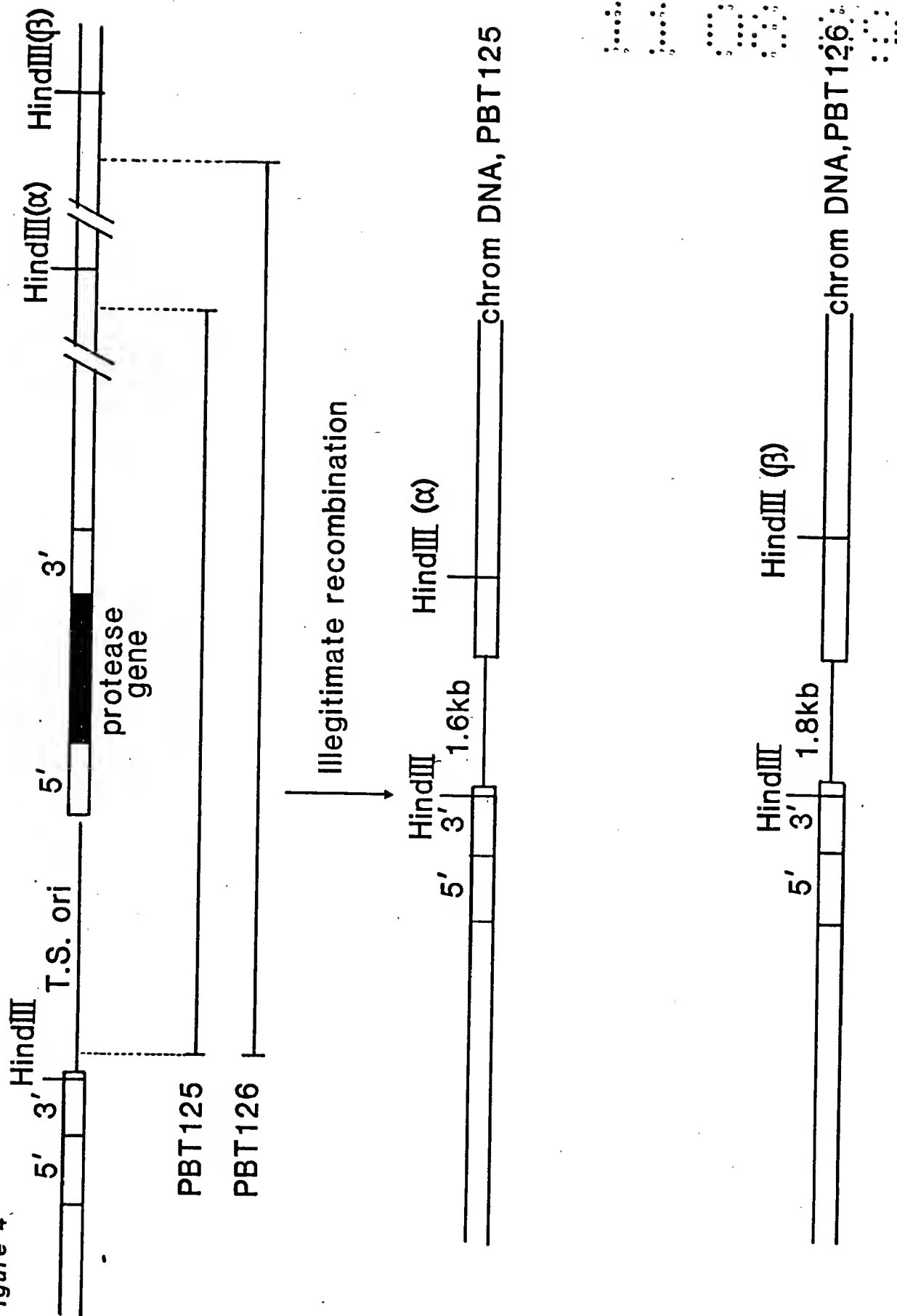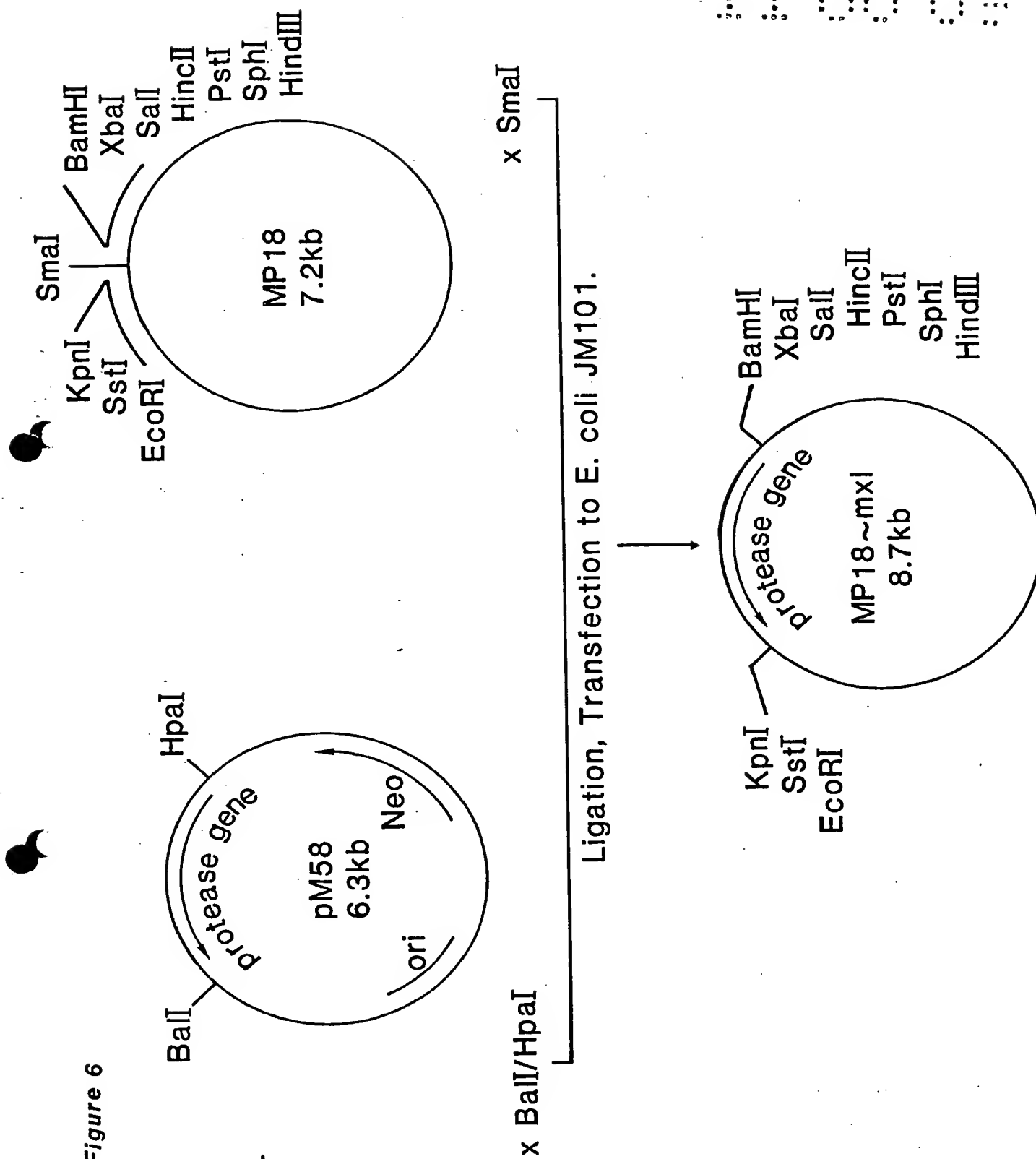
Figure 6

figure 7

Sequence of pBHA-1

```
                    30                              60
AATTCACCTC GAAAGCAAGC TGATAAACCG ATACAATTAA AGGCTCCTTT TGGAGCCTTT

                    90                             120
TTTTTTGGAG ATTTTCAACG TGAAAAAATT ATTATTCGCA ATTCCAAGCT AATTCACCTC

                    150                            180
GAAAGCAAGC TGATAAACCG ATACAATTAA AGGCTCCTTT TGGAGCCTTT TTTTTTGGAG

                    210                            240
ATTTTCAACG TGAAAAAATT ATTATTCGCA ATTCCAAGCT CTGCCTCGCG CGTTTCGGTG

                    270                            300
ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTCACAGCT TGTCTGTAAG

                    330                            360
CGGATGCAGA TCACGCGCCC TGTAGCGGCG CATTAAGCGC GGCGGGTGTG GTGGTTACGC

                    390                            420
GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCGC TCCTTTCGCT TTCTTCCCTT

                    450                            480
CCTTTCTCGC CACGTTCGCC GGCTTTCCCC GTCAAGCTCT AAATCGGGGG CTCCCTTTAG

                    510                            540
GGTTCCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG GGTGATGGTT

                    570                            600
CACGTAGTGG GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG GAGTCCACGT

                    630                            660
TCTTTAATAG TGGACTCTTG TTCCAAACTG GAACAACACT CAACCCTATC TCGGTCTATT

                    690                            720
CTTTTGATTT ATAAGGGATT TTGCCGATTT CGGCCTATTG GTTAAAAAAT GAGCTGATTT

                    750                            780
AACAAAAATT TAACGCGAAT TTTAACAAAA TATTAACGTT TACAATTTGA TCTGCGCTCG

                    810                            840
GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA

                    870                            900
GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC

                    930                            960
CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC

                    990                            1020
AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG

                    1050                           1080
TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC
```

```
                  1110                               1140
CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC AATGCTCACG CTGTAGGTAT

                  1170                               1200
CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG

                  1230                               1260
CCCGACCGCT GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC

                  1290                               1320
TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT

                  1350                               1380
GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGGAC AGTATTTGGT

                  1410                               1440
ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC

                  1470                               1500
AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA

                  1530                               1560
AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC

                  1590                               1620
GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC

                  1650                               1680
CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA AACTTGGTCT

                  1710                               1740
GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA

                  1770                               1800
TCCATAGTTG CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT

                  1830                               1860
GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA

                  1890                               1920
ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC

                  1950                               1980
ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAATAGTTTG

                  2010                               2040
CGCAACGTTG TTGCCATTGC TGCAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT

                  2070                               2100
TCATTCAGCT CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA

                  2130                               2160
AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA

                  2190                               2220
TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC
```

```
                       2250                              2380
TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT GCGGCGACCG

                       2310                              2340
AGTTGCTCTT GCCCGGCGTC AACACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA

                       2370                              2400
GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG

                       2430                              2460
AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC

                       2490                              2520
ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG

                       2550                              2580
GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG AAGCAGACAG

                       2610                              2640
TTTTATTGTT CATGATGATA TATTTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA

                       2670                              2700
CACAACGTGG CTTTGTTGAA TAAATCGAAC TTTTGCTGAG TTGACTCCCC GCGCGCGATG

                       2730                              2760
GGTCGAATTT GCTTTCGAAA AAAAAGCCCG CTCATTAGGC GGGCTAAAAA AAAGCCCGCT

                       2790                              2820
CATTAGGCGG GCTCGAATTT CTGCCATTCA TCCGCTTATT ATCACTTATT CAGGCGTAGC

                       2850                              2880
AACCAGGCGT TTAAGGGCAC CAATAACTGC CTTAAAAAAA TTACGCCCCG CCCTGCCACT

                       2910                              2940
CATCGCAGTA CTGTTGTAAT TCATTAAGCA TTCTGCCGAC ATGGAAGCCA TCACAGACGG

                       2970                              3000
CATGATGAAC CTGAATCGCC AGCGGCATCA GCACCTTGTC GCCTTGCGTA TAATATTTGC

                       3030                              3060
CCATAGTGAA AACGGGGGCG AAGAAGTTGT CCATATTCGC CACGTTTAAA TCAAAACTGG

                       3090                              3120
TGAAACTCAC CCAGGGATTG GCTGAGACGA AAAACATATT CTCAATAAAC CCTTTAGGGA

                       3150                              3180
AATAGGCCAG GTTTTCACCG TAACACGCCA CATCTTGCGA ATATATGTGT AGAAACTGCC

                       3210                              3240
GGAAATCGTC GTGGTATTCA CTCCAGAGCG ATGAAAACGT TTCAGTTTGC TCATGGAAAA

                       3270                              3300
CGGTGTAACA AGGGTGAACA CTATCCCATA TCACCAGCTC ACCGTCTTTC ATTGCCATAC

                       3330                              3360
GAAATTCCGG ATGAGCATTC ATCAGGCGGG CAAGAATGTG AATAAAGGCC GGATAAAACT
```

```
                      3390                                         3420
TGTGCTTATT TTTCTTTACG GTCTTTAAAA AGGCCGTAAT ATCCAGCTAA ACGGTCTGGT

                      3450                                         3480
TATAGGTACA TTGAGCAACT GACTGAAATG CCTCAAAATG TTCTTTACGA TGCCATTGGG

                      3510                                         3540
ATATATCAAC GGTGGTATAT CCAGTGATTT TTTTCTCCAT TTTAGCTTCC TTAGCTCCTG

                      3570                                         3600
AAAATCTCGA TAACTCAAAA AATACGCCCG GTAGTGATCT TATTTCATTA TGGTGAAAGT

                      3630                                         3660
TGGAACCTCT TACGTGCCGA TCAACGTCTC ATTTTCGCCA AAAGTTGGCC CAGGGCTTCC

                      3690                                         3720
CGGTATCAAC AGGGACACCA GGATTTATTT ATTCTGCGAA GTGATCTTCC GTCACAGGTA

                      3750                                         3780
TTTATTCGAA GACGAAAGGG CATCGCGCGC GGGGAATTCC CGGGAGAGCT CGATATCGCA

                      3810                                         3840
TGCGGTACCT CTAGAAGAAG CTTGGAGACA AGGTAAAGGA TAAAACAGCA CAATTCCAAG

                      3870                                         3900
AAAAACACGA TTTAGAACCT AAAAAGAACG AATTTGAACT AACTCATAAC CGAGAGGTAA

                      3930                                         3960
AAAAAGAACG AAGTCGAGAT CAGGGAATGA GTTTATAAAA TAAAAAAAGC ACCTGAAAAG

                      3990                                         4020
GTGTCTTTTT TTGATGGTTT TGAACTTGTT CTTTCTTATC TTGATACATA TAGAAATAAC

                      4050                                         4080
GTCATTTTTA TTTTAGTTGC TGAAAGGTGC GTTGAAGTGT TGGTATGTAT GTGTTTTAAA

                      4110                                         4140
GTATTGAAAA CCCTTAAAAT TGGTTGCACA GAAAAACCCC ATCTGTTAAA GTTATAAGTG

                      4170                                         4200
ACTAAACAAA TAACTAAATA GATGGGGGTT TCTTTTAATA TTATGTGTCC TAATAGTAGC

                      4230                                         4260
ATTTATTCAG ATGAAAAATC AAGGGTTTTA GTGGACAAGA CAAAAAGTGG AAAAGTGAGA

                      4290                                         4320
CCATGGAGAG AAAAGAAAAT CGCTAATGTT GATTACTTTG AACTTCTGCA TATTCTTGAA

                      4350                                         4380
TTTAAAAAGG CTGAAAGAGT AAAAGATTGT GCTGAAATAT TAGAGTATAA ACAAAATCGT

                      4410                                         4440
GAAACAGGCG AAAGAAAGTT GTATCGAGTG TGGTTTTGTA AATCCAGGCT TTGTCCAATG

                      4470                                         4500
TGCAACTGGA GGAGAGCAAT GAAACATGGC ATTCAGTCAC AAAAGGTTGT TGCTGAAGTT
```

```
                        4530                           4560
ATTAAACAAA AGCCAACAGT TCGTTGGTTG TTTCTCACAT TAACAGTTAA AAATGTTTAT

                        4590                           4620
GATGGCGAAG AATTAAATAA GAGTTTGTCA GATATGGCTC AAGGATTTCG CCGAATGATG

                        4650                           4680
CAATATAAAA AAATTAATAA AAATCTTGTT GGTTTTATGC GTGCAACGGA AGTGACAATA

                        4710                           4740
AATAATAAAG ATAATTCTTA TAATCAGCAC ATGCATGTAT TGGTATGTGT GGAACCAACT

                        4770                           4800
TATTTTAAGA ATACAGAAAA CTACGTGAAT CAAAAACAAT GGATTCAATT TTGGAAAAAG

                        4830                           4860
GCAATGAAAT TAGACTATGA TCCAAATGTA AAAGTTCAAA TGATTCGACC GAAAAATAAA

                        4890                           4920
TATAAATCGG ATATACAATC GGCAATTGAC GAAACTGCAA AATATCCTGT AAAGGATACG

                        4950                           4980
GATTTTATGA CCGATGATGA AGAAAAGAAT TTGAAACGTT TGTCTGATTT GGAGGAAGGT

                        5010                           5040
TTACACCGTA AAAGGTTAAT CTCCTATGGT GGTTTGTTAA AAGAAATACA TAAAAAATTA

                        5070                           5100
AACCTTGATG ACACAGAAGA AGGCGATTTG ATTCATACAG ATGATGACGA AAAAGCCGAT

                        5130                           5160
GAAGATGGAT TTTCTATTAT TGCAATGTGG AATTGGGAAC GGAAAAATTA TTTTATTAAA

                        5190                           5220
GAGTAGTTCA ACAAACGGGC CAGTTTGTTG AAGATTAGAT GCTATAATTG TTATTAAAAG

                        5250                           5280
GATTGAAGGA TGCTTAGGAA GACGAGTTAT TAATAGCTGA ATAAGAACGG TGCTCTCCAA

                        5310                           5340
ATATTCTTAT TTAGAAAAGC AAATCTAAAA TTATCTGAAA AGGGAATGAG AATAGTGAAT

                        5370                           5400
GGACCAATAA TAATGACTAG AGAAGAAAGA ATGAAGATTG TTCATGAAAT TAAGGAACGA

                        5430                           5460
ATATTGGATA AATATGGGGA TGATGTTAAG GCTATTGGTG TTTATGGCTC TCTTGGTCGT

                        5490                           5520
CAGACTGATG GGCCCTATTC GGATATTGAG ATGATGTGTG TCATGTCAAC AGAGGAAGCA

                        5550                           5580
GAGTTCAGCC ATGAATGGAC AACCGGTGAG TGGAAGGTGG AAGTGAATTT TGATAGCGAA

                        5610                           5640
GAGATTCTAC TAGATTATGC ATCTCAGGTG GAATCAGATT GGCCGCTTAC ACATGGTCAA
```

```
                    5670                                        5700
TTTTTCTCTA TTTTGCCGAT TTATGATTCA GGTGGATACT TAGAGAAAGT GTATCAAACT

                    5730                              5760
GCTAAATCGG TAGAAGCCCA AACGTTCCAC GATGCGATTT GTGCCCTTAT CGTAGAAGAG

                    5790                              5820
CTGTTTGAAT ATGCAGGCAA ATGGCGTAAT ATTCGTGTGC AAGGACCGAC AACATTTCTA

                    5850                              5880
CCATCCTTGA CTGTACAGGT AGCAATGGCA GGTGCCATGT TGATTGGTCT GCATCATCGC

                    5910                              5940
ATCTGTTATA CGACGAGCGC TTCGGTCTTA ACTGAAGCAG TTAAGCAATC AGATCTTCCT

                    5970                              6000
TCAGGTTATG ACCATCTGTG CCAGTTCGTA ATGTCTGGTC AACTTTCCGA CTCTGAGAAA

                    6030                              6060
CTTCTGGAAT CGCTAGAGAA TTTCTGGAAT GGGATTCAGG AGTGGACAGA ACGACACGGA

                    6090                              6120
TATATAGTGG ATGTGTCAAA ACGCATACCA TTTTGAACGA TGACCTCTAA TAATTGTTAA

                    6150                              6180
TCATGTTGGT TACGTATTTA TTAACTTCTC CTAGTATTAG TAATTATCAT GGCTGTCATG

                    6210                              6240
GCGCATTAAC GGAATAAAGG GTGTGCTTAA ATCGGGCCAT TTTGCGTAAT AAGAAAAAGG

                    6270                              6300
ATTAATTATG AGCGAATTGA ATTAATAATA AGGTAATAGA TTTACATTAG AAAATGAAAG

                    6330                              6360
GGGATTTTAT GCGTGAGAAT GTTACAGTCT ATCCCGGCAT TGCCAGTCGG GGATATTAAA

                    6390                              6420
AAGAGTATAG GTTTTTATTG CGATAAACTA GGTTTCACTT TGGTTCACCA TGAAGATGGA

                    6450                              6480
TTCGCAGTTC TAATGTGTAA TGAGGTTCGG ATTCATCTAT GGGAGGCAAG TGATGAAGGC

                    6510                              6540
TGGCGCTCTC GTAGTAATGA TTCACCGGTT TGTACAGGTG CGGAGTCGTT TATTGCTGGT

                    6570                              6600
ACTGCTAGTT GCCGCATTGA AGTAGAGGGA ATTGATGAAT TATATCAACA TATTAAGCCT

                    6630                              6660
TTGGGCATTT TGCACCCCAA TACATCATTA AAAGATCAGT GGTGGGATGA ACGAGACTTT

                    6690                              6720
GCAGTAATTG ATCCCGACAA CAATTTGATT AGCTTTTTTC AACAAATAAA AAGCTAAAAT

                    6750                              6780
CTATTATTAA TCTGTTCAGC AATCGGGCGC GATTGCTGAA TAAAAGATAC GAGAGACCTC
```

```
                          6810                          6840
TCTTGTATCT TTTTTATTTT GAGTGGTTTT GTCCGTTACA CTAGAAAACC GAAAGACAAT

                          6870                          6900
AAAAATTTTA TTCTTGCTGA GTCTGGCTTT CGGTAAGCTA GACAAAACGG ACAAAATAAA

                          6930                          6960
AATTGGCAAG GGTTTAAAGG TGGAGATTTT TTGAGTGATC TTCTCAAAAA ATACTACCTG

                          6990                          7020
TCCCTTGCTG ATTTTTAAAC GAGCACGAGA GCAAAACCCC CCTTTGCTGA GGTGGCAGAG

                          7050                          7080
GGCAGGTTTT TTTGTTTCTT TTTTCTCGTA AAAAAAAGAA AGGTCTTAAA GGTTTTATGG

                          7110                          7140
TTTTGGTCGG CACTGCCGAC AGCCTCGCAG GACACACACT TTATGAATAT AAAGTATAGT

                          7170                          7200
GTGTTATACT TTACTTGGAA GTGGTTGCCG GAAAGAGCGA AAATGCCTCA CATTTGTGCC

                          7230                          7260
ACCTAAAAAG GAGCGATTTA CATATGAGTT ATGCAGTTTG TAGAATGCAA AAAGTGAAAT

                          7290                          7320
CAGGGGGATC CTCTAGAGTC GAGCTCAAGC TAGCTTGGTA CGTACCAGAT CTGAGATCAC

GCGTTCTAGA GGTCGA
```
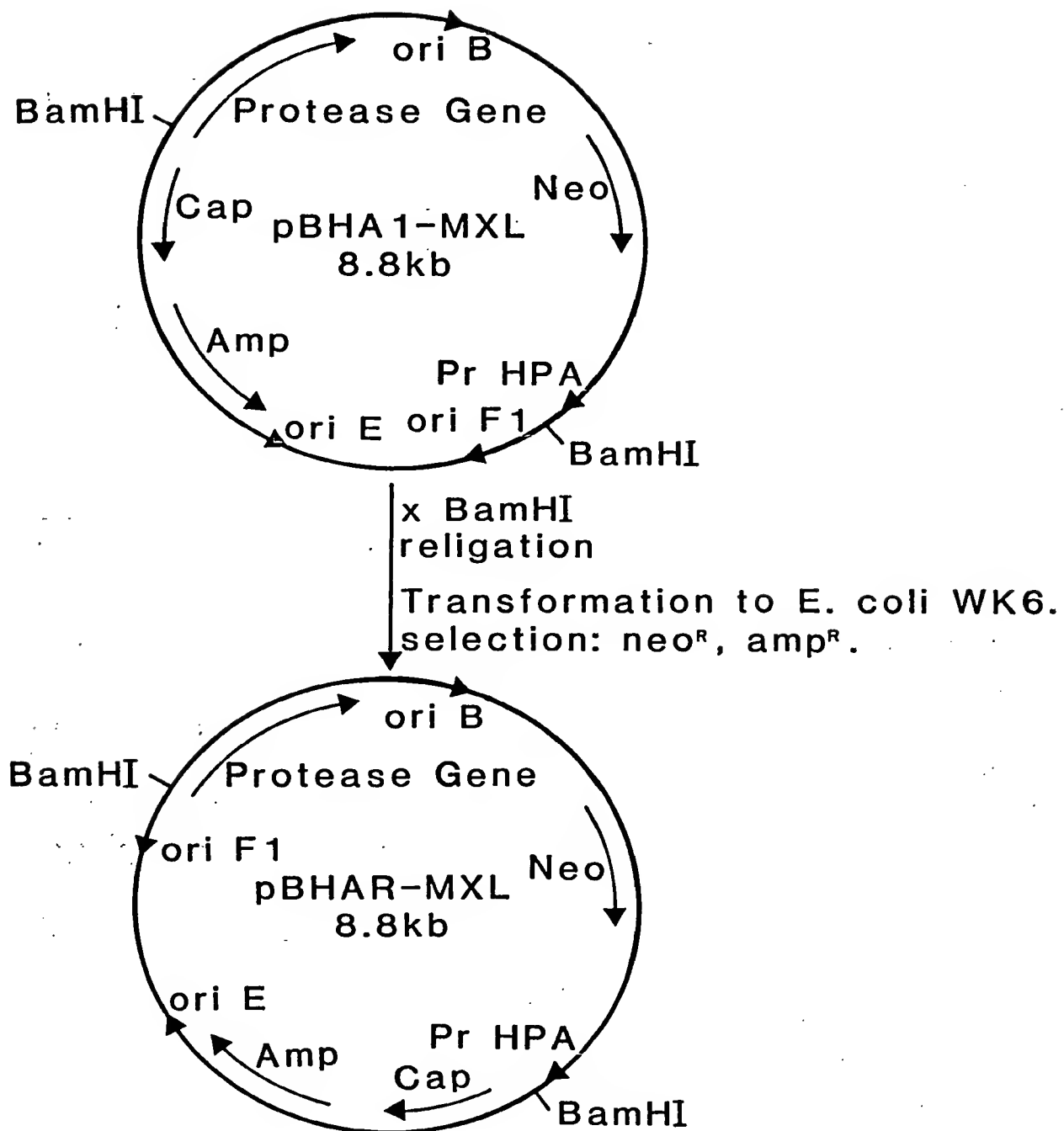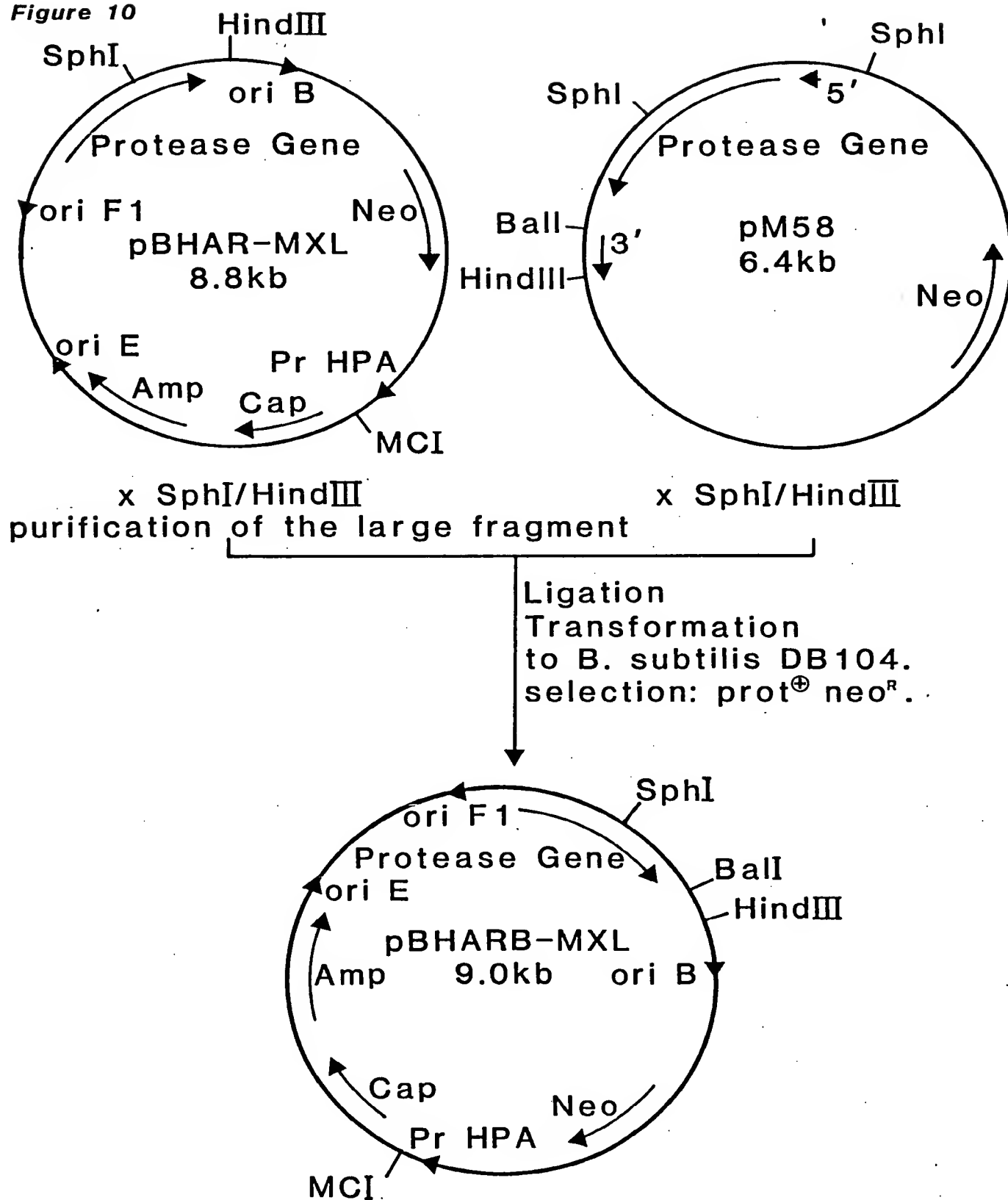
Figure 8

Figure 9

11 06 89

Figure 10
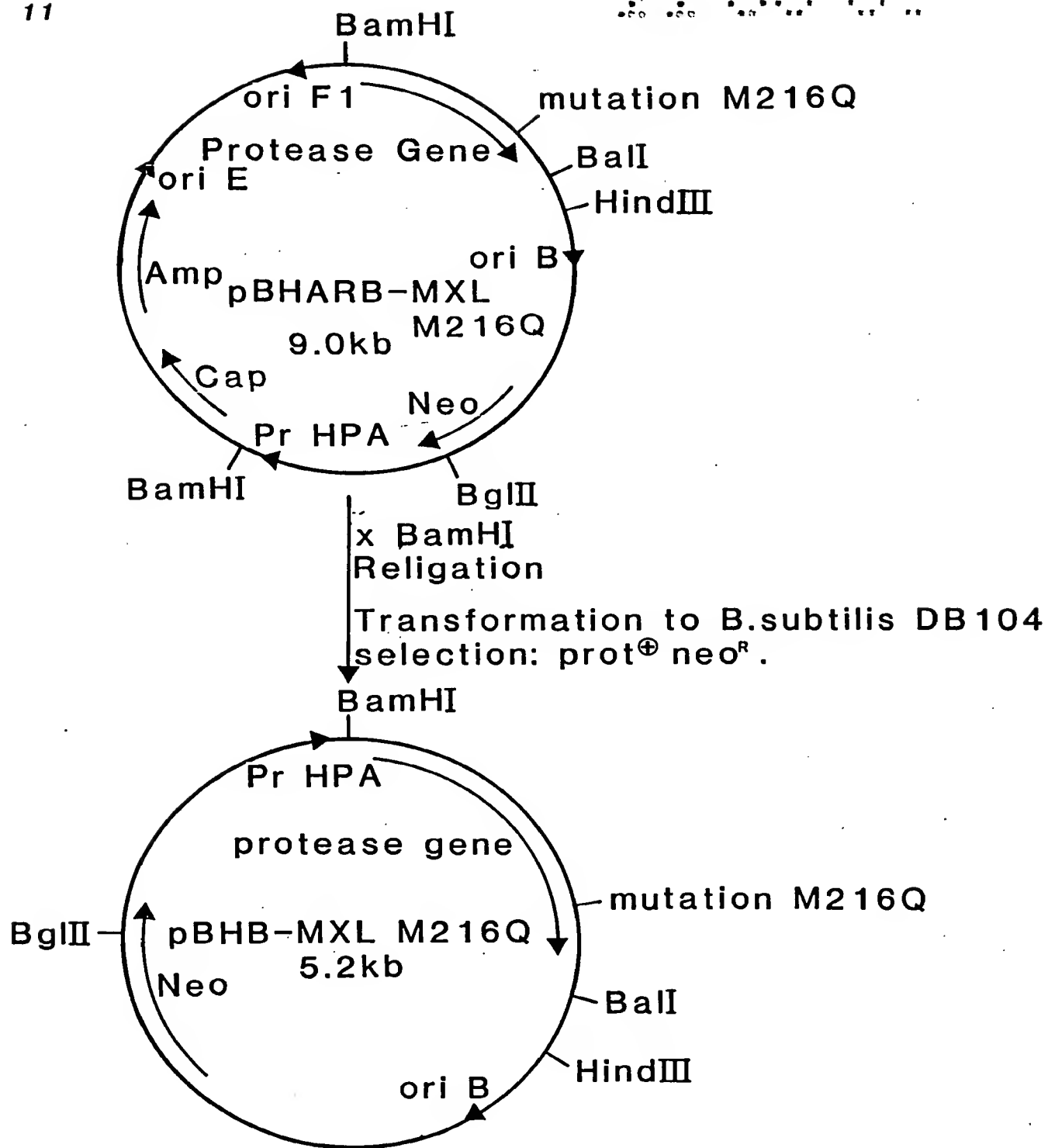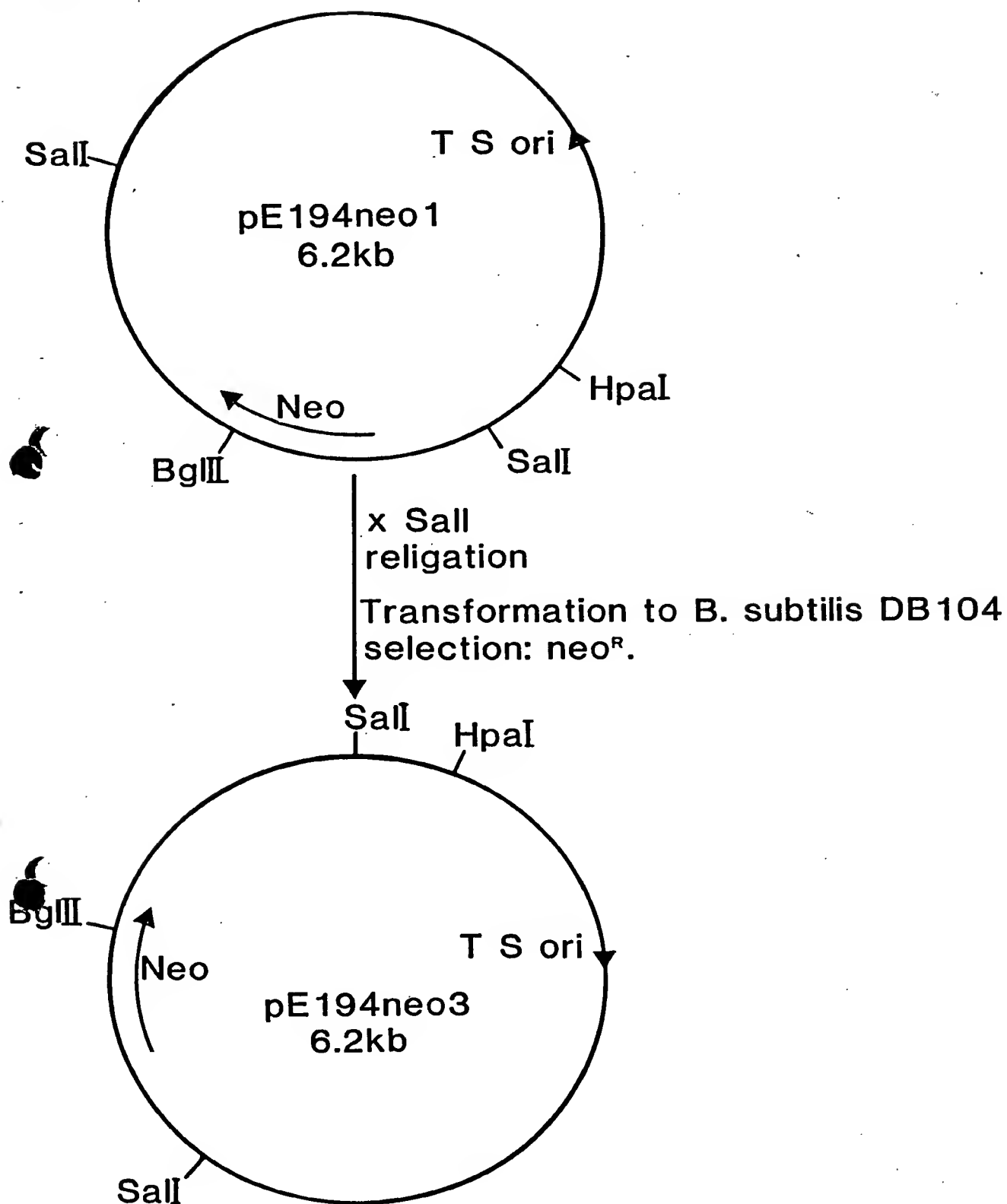
Figur 11

Figure 12

Figure 13

Figure 14